

# Robert Calef

Cambridge, MA | 510.409.4280 | rcalef@mit.edu

---

## **Education**

### **Ph.D., EECS**

Advised by Prof. Marinka Zitnik and Prof. Manolis Kellis

Machine learning for biological discovery

MIT

August 2023-Present

- Protein representation learning and multi-modal large language models fusing text and protein understanding to drive insight in understudied parts of the human proteome

### **Bachelor of Science in Bioinformatics, *summa cum laude***

Minor in Mathematics

University of California, Santa Cruz

June 2015

## **Professional Experience**

### **Staff Machine Learning Scientist**

August 2021 - August 2023

### **Senior Scientist, Bioinformatics**

September 2020 - August 2021

### **Bioinformatics Scientist 2**

December 2017 - September 2020

GRAIL, Inc., Menlo Park, CA

- Key contributor to all facets of GRAIL's machine learning pipeline for training and evaluating a cancer detection classifier using methylation information from hundreds of terabytes of data from plasma cell-free DNA sequencing. Including:
  - Scalable pipelines for cleaning and preprocessing of DNA sequencing data for tens of thousands of patient samples
  - Distributed computing methods for feature engineering and selection to reduce data per training example from hundreds of gigabytes to ~30k features
  - Classification model training using a custom two-stage classification architecture for presence of cancer and tissue-of-origin. Implementations of ADAM and dropout for MLP training, followed by score calibration and rule-based label calling to capture uncertainty in final tissue prediction
  - All software described above was written according to best practices using source control, testing, and code review, and is now included in GRAIL's commercial cancer test, Galleri.
- Led a cross-functional research initiative with a \$300k budget to characterize cancer signal availability in plasma samples from clinically diagnosed patients. Designed study and coordinated execution with operations and wet lab peers. Developed all analysis code including a novel somatic nucleotide variant calling pipeline, targeted sequencing panel design, and Bayesian models for tumor fraction estimation from noisy data.
- Prototyped cancer classification features using a sample's sequencing coverage and viral DNA fragments in plasma. Exploration of human papillomavirus (HPV) signals created better understanding of the effects of HPV infection on the human methylome, and informed improvements to tissue-of-origin classification accuracy.
- Organized a summer internship program hosting three current Ph.D. students for three months. Organization included arranging on-site visit, speaker series from other departments in the company, and happy hours with other interns and team members.
  - Mentored one intern for a summer project focused on analyzing the intersection of transcription factor binding sites with regions of differential methylation in multiple cancers

### **Intern → Bioinformatics Analyst → Bioinformatics Scientist**

April 2015 - November 2017

Dovetail Genomics LLC, Santa Cruz, CA

- Developed novel algorithms and runtime optimizations as one of three primary developers of HiRise, a statistical model and software pipeline for genome assembly scaffolding using a novel DNA

sequencing data type. HiRise has been used to generate hundreds of highly contiguous and validated genome assemblies.

- Drove over one hundred customer genome scaffolding projects as informatics lead. Directly communicated with external researchers to answer technical inquiries and support downstream analyses.

## **Research Experience**

### **Undergraduate Researcher and Systems Administrator**

March 2014 - June 2015

*UCSC Paleogenomics Lab, Santa Cruz, CA*

- Designed and implemented a program using multiple genome alignment data to identify genomic regions that are informative for determining archosaurian ancestry in sequencing reads from a fossilized sample.

## **Publications**

- Jamshidi A, Liu MC, Klein EA, ..., **Calef R**, Aravanis AM, Seiden MV, Swanton C. Evaluation of cell-free DNA approaches for multi-cancer early detection. *Cancer Cell* (2022).
- Liu MC, Oxnard GR, Klein EA, Swanton C, Seiden MV, ... , **Calef R**, ..., Curtis C, Berry DA. Sensitive and specific multi-cancer detection and localization using methylation signatures in cell-free DNA. *Annals of Oncology* **31** (6), 745-759 (2020).
- Simakov O, Marlétaz F, Yue JX, O'Connell BL, Jenkins J, Brandt A, **Calef R**, Tung CH, Huang TK, Schmutz J, Satoh N, Yu JK, Putnam NH, Green RE, Rokhsar DS. Deeply conserved synteny resolves early events in vertebrate evolution. *Nature Ecology & Evolution* **4** (6), 820–830 (2020).
- Putnam NH, O'Connell BL, Stites JC, Rice BJ, Blanchette M, **Calef R**, Troll CJ, Fields A, Hartley PD, Sugnet CW, Haussler D, Rokhsar DS, Green RE. Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. *Genome Res* **26** (3), 342–350 (2016).
- **Calef R**, Gross S, Fields A, Zhao M, Betts C, Burkhardt D, Venn O. Multi-cancer measurement of low tumor fraction in cell-free DNA using patient-specific targeted sequencing. **Pending submission to *Clinical Cancer Research*.**

## **Congress Presentations**

- **Calef R**, Venn O, Maher MC, Beausang JF, Hubbell E, Patel A, Fields AP, Bredno J, Jamshidi A, Aravanis AM. HPV-driven cancers show distinct methylation signatures in cell-free DNA. American Association for Cancer Research Annual Meeting 2020. Poster 2114.
- Melton C, Freese P, Bagaria S, Hubbell E, Filippova D, **Calef R**, Maher MC, Nicula V, Gross S, Venn O, Valouev A, Aravanis AM, Jamshidi A. Optimized Early Cancer Detection from Whole-Genome Sequencing of Cell-Free DNA. American Society of Human Genetics Annual Meeting 2019. Poster 2412.
- Filippova D, Larson MH, Maher MC, **Calef R**, Pimentel M, Zhou Y, Newman J, Gross S, Nicula V, Liu TC, Yakym C, Berman J, Aravanis AM, Jamshidi A. The Circulating Cell-free Genome Atlas (CCGA) study: size selection of cell-free DNA fragments. American Society of Clinical Oncology Annual Meeting 2019. Poster 3103.

## **Patent Submissions**

- Venn O, Freese P, Gross S, **Calef R**, Jamshidi A. 2021. Conditional tissue of origin return for localization accuracy. U.S. Patent US17/714,062, filed April 6, 2021. Patent pending.
- Sakarya O, Yakym C, Singh P, **Calef R**, Huang R. 2020. Detecting cross-contamination in sequencing data. U.S. Patent 17/478,856, filed September 18, 2020. Patent pending.
- **Calef R**, Maher MC, Beausang JF, Bredno J, Venn O, Fields AP, Jamshidi A. 2020. U.S. Patent 17/350,511, filed June 6, 2020. Patent pending.
- Xiang J, **Calef R**. Systems and methods for estimating cell source fractions using methylation information. 2019. U.S. Patent 17/127,813, filed December 18, 2019. Patent pending.

- Liu Q, Venn O, Gross S, **Calef R**. Cancer classification with tissue of origin thresholding. U.S. Patent 17/066,863, filed October 11, 2019, Patent pending.
- Fields AP, Beausang JF, Venn O, Jamshidi A, Maher MC, Liu Q, Schellenberger J, Newman J, **Calef R**, Gross S. Model-based featurization and classification. U.S. Patent 15/931,022, filed May 13, 2019. Patent pending.
- Filippova D, Larson MH, Maher MC, Pimentel M, **Calef R**. Systems and methods for enriching for cancer-derived fragments using fragment size. U.S. Patent 16/816,918, filed March 13, 2019. Patent pending.
- Maher MC, Valouev A, Filippova D, Nicula V, Jagadeesh K, Venn O, Gross S, Beausang JF, **Calef R**. Systems and methods for classifying patients with respect to multiple cancer classes. U.S. Patent 16/709,537, filed December 10, 2018. Patent pending.
- Nicula V, Valouev A, Filippova D, Larson MH, Maher MC, Pimentel M, **Calef R**, Melton C. Convolutional neural network systems and methods for data classification. U.S. Patent 16/428,575, filed June 1, 2018, **and issued** October 25, 2022.
- Filippova D, Valouev A, Nicula V, Jagadeesh K, Maher MC, Larson MH, Pimentel M, **Calef R**. Method and system for selecting, managing, and analyzing data of high dimensionality. U.S. Patent 16/352,739, filed March 13, 2018. Patent pending.

### **Awards**

- Hierarchical Systems Research Foundation Scholarship, 2013-14.
- Baskin Endowed Scholarship, 2013-14.
- C. Fiore Chemistry Prize, 2013-14.
- Jack Baskin Engineering Scholarship, 2012-13.