

---

## BIOGRAPHICAL SKETCH

Provide the following information for the Senior/key personnel and other significant contributors in the order listed on Form Page 2.

Follow this format for each person. **DO NOT EXCEED FIVE PAGES.**

---

NAME: Fertig, Elana

eRA COMMONS USER NAME (credential, e.g., agency login): efertig1

POSITION TITLE: Assistant Director of Quantitative Sciences, Associate Professor of Oncology and Applied Mathematics and Statistics

---

EDUCATION/TRAINING (*Begin with baccalaureate or other initial professional education, such as nursing, include postdoctoral training and residency training if applicable.*)

---

INSTITUTION AND LOCATION	DEGREE (if applicable)	MM/YY	FIELD OF STUDY
Brandeis University, Waltham, MA	BS	2003	Mathematics / Physics
University of Maryland, College Park, MD	MS	2005	Applied Mathematics
University of Maryland, College Park, MD	PhD	2007	Applied Mathematics
Johns Hopkins University, Baltimore, MD	Postdoc	2010	Cancer Bioinformatics and Systems Biology

### A. Personal Statement

This proposal aims to develop latent space methods to infer composite biological processes from high-dimensional scRNA-seq data in the Human Cell Atlas. To this project, I bring expertise in algorithm development and applied bioinformatics analysis. I was lead developer of the CoGAPS Markov chain Monte Carlo (MCMC) pattern identification algorithm. CoGAPS analysis of genomics data infers cellular signaling processes in cancer subtypes, perturbation experiments, and developmental time-course. My research has been featured in over 50 peer-reviewed publications and recognized in NCI R01, U01, and Collaborative Computational Tools for the Human Cell Atlas grants. In the latter project, we developed novel extensions of CoGAPS for scRNA-seq analysis in the developing retina and a new transfer learning approach to relate these patterns across measurement technologies, organisms, and developmental time. In these projects, I have collaborated with numerous cross-disciplinary teams including several members of the latent spaces team and other groups in the Collaborative Computational Tools for the Human Cell Atlas group. Accordingly, I frequently communicate with collaborators to ensure progress and the translational relevance of the results from genomics data analysis. In summary, I have developed a successful record of novel computational analysis algorithms and bioinformatics analysis that have prepared me to develop latent space methods for the Human Cell Atlas in the proposed project.

### B. Positions and Honors

#### Positions and Employment

2007 – 2008	Analyst, Metron, Inc., Reston, VA,
2008 – 2010	Research Fellow, Oncology Biostatistics, Johns Hopkins University School of Medicine
2010 – 2013	Instructor, Oncology Biostatistics, Johns Hopkins University School of Medicine
2013 – 2018	Assistant Professor, Oncology Biostatistics, Johns Hopkins University School of Medicine
2018 – Present	Assistant Director of the Research Program of Quantitative Sciences, Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University School of Medicine
2018 – Present	Associate Professor, Oncology Biostatistics, Johns Hopkins University School of Medicine
2018 – Present	Associate Professor, Applied Mathematics and Statistics, Johns Hopkins University

#### Honors

2012	IEEE BIBM 2012 Best Paper Award
2013	Awarded the Helen Masenhimer Fellowship through the Cleveland Foundation
2013	Led winning team on the HPN-DREAM Breast Cancer Network Inference Challenge 2A
2013 – 2014	Accepted participant in Emerging Women Leadership Program, JHU, Baltimore MD.
2014	Winner of the Summer Science Program New Curriculum Prize.
2016	PI of JHU Discovery Award from the provost office for program building.

2017 PI of JHU Catalyst Award from the provost office for early-career faculty.  
 2017 PI of Synergy Award for innovative multi-disciplinary research in the JHUSOM.

### C. Contribution to Science

1. Pattern detection in high-throughput genomics data with matrix factorization: My computational biology research is characterized by the development of biologically-driven genomics algorithms. In total, the Bioconductor packages including these tools were downloaded over 4,500 times from 2,449 unique IPs in 2017. I led development of a novel pattern identification algorithm, Coordinated Gene Activity in Pattern Sets (CoGAPS). CoGAPS is a smooth-sparse, Bayesian non-negative matrix factorization algorithm implemented in an R/Bioconductor package by the same name. This algorithm identifies sets of genes, called meta-pathways, with concurrent changes in high-throughput data. CoGAPS also provides a continuous measure of the extent to which each meta-pathway is active in specific samples. This meta-pathway activity can distinguish sample subtypes or dynamics of biological processes. For example, we found that CoGAPS meta-pathway activity reflected the activation of transcription factor networks when gastrointestinal stromal tumors respond to targeted therapeutics more accurately than other pattern identification algorithms. We applied CoGAPS in numerous additional applications, including notably for precision medicine in cancer as described in 1 above. CoGAPS algorithm is widely applicable, including to data integration to learn tumor subtypes and time course analysis of transcriptional changes in brain development. Recently, we developed a new genome-wide approach (a) that enables CoGAPS to distinguish tissue-specific gene expression signatures from healthy human tissues in GTEx and developed a standardized language for results from matrix factorization algorithms (b). Our recent extension of CoGAPS to scRNA-seq data as part of the Collaborative Computational Tools for the Human Cell Atlas by the Chan Zuckerberg Initiative award enabled inference of cell types, transitions, and sex-specific differences during retinal development (c). Moreover, we demonstrated that these patterns related to complex biological processes shared across data modalities, developmental time, and organisms with a novel transfer learning approach ProjectR (d). Altogether, these results demonstrate the power of scCoGAPS as a latent space tool for the Human Cell Atlas.
  - a. Stein-O'Brien GL, Carey JL, Lee W-S, Considine M, Favorov AV, Flam E, Guo T, Li S, Marchionni L, Sherman T, Sivy S, Gaykalova DA, McKay RD, Ochs MF, Colantuoni C, and **Fertig EJ**. (2017) PatternMarkers & GWCoGAPS for novel data-driven biomarkers via whole transcriptome NMF. *Bioinformatics*. **33**:1892-1894. PMID 28174896. PMCID PMC5860188.
  - b. Stein-O'Brien GL, Arora R, Culhane AC, Favorov AV, Garmire LX, Greene CS, Goff LA, Li Y, Ngom A, Ochs MF, Xu Y, and **Fertig EJ**. (2018) Enter the Matrix: Factorization uncovers knowledge from omics. *Trends in Genetics*. **34**:780-805. PMID 30143323.
  - c. Clark B, Stein-O'Brien GL, Shiao F, Cannon G, Davis E, Sherman T, Rajaii F, James-Esposito R, Gronostajski R, **Fertig E**, Goff L, and Blackshaw S. (2018) Comprehensive analysis of retinal development at single cell resolution identifies NFI factors as essential for mitotic exit and specification of late-born cells. *Bioarxiv*. <https://doi.org/10.1101/378950>
  - d. Stein-O'Brien GL, Clark B, Sherman T, Zibetti C, Hu Q, Sealfon R, Liu S, Qian J, Colantuoni C, Blackshaw S, Goff LA, and **Fertig EJ**. (2018) Decomposing cell identity for transfer learning across cellular measurements, platforms, tissues, and species. *Bioarxiv*. <https://doi.org/10.1101/395004>
2. Bioinformatics methods for high-throughput genomics analysis and tumor heterogeneity: In addition to CoGAPS, my lab has developed and contributed to numerous novel algorithms for genomics analysis in cancer. One example includes the GSReg package developed in my lab to model tumor heterogeneity. Specifically, this package includes a multivariate non-parametric statistic called Expression Variation Analysis (EVA, a) to quantify pathway dysregulation between tumor samples from distinct subtypes. The statistics in EVA were extended in new algorithms to model differential heterogeneity of splice variant usage in tumors (SEVA, b) and chromatin regulation of gene expression (c). Together, these applications mark EVA as a robust suite of tools to model transcriptional heterogeneity in cancer. My lab has contributed additional algorithms spanning problems in data preprocessing (d), genome-wide data integration, and machine learning which are all disseminated to the public as software packages.
  - a. Afsari B, Geman D, and **Fertig EJ**. (2014) Learning dysregulated pathways in cancers from differential variability analysis. *Cancer Informatics*. **13**:61-67. PMID 25392694. PMCID PMC4218688.

- b. Afsari B, Guo T, Considine M, Florea L, Kagohara LT, Stein-O'Brien GL, Kelley D, Flam E, Zambo KD, Ha PK, Geman D, Ochs MF, Califano JA, Gaykalova DA, Favorov AV, and **Fertig EJ**. (2018) Splice Expression Variation Analysis (SEVA) for Inter-tumor Heterogeneity of Gene Isoform Usage in Cancer. *Bioinformatics: Epub ahead of print*. PMID: 29342249.
    - c. Kelley DZ, Flam EL, Izumchenko E, Danilova LV, Wulf HA, Guo T, Singman DA, Bahman A, Skaist A, Considine M, Stavrovskaya E, Bishop JA, Westra WH, Khan Z, Koch WM, Sidransky D, Wheelan S, Favorov AV, Califano JA, **Fertig EJ\***, Gaykalova DA\*. (2017) Integrated analysis of whole-genome ChIP-Seq and RNA-Sequencing data of primary tumor samples associates HPV integration sites with open chromatin marks. *Cancer Research*. PMID: 28947419. PMCID in process. \*Equal Contribution.
    - d. Parker HS, Leek JT, Favorov AV, Considine M, Xia X, Chavan S, Chung CH, and **Fertig EJ** (2014) Preserving biological heterogeneity with a permuted surrogate variable analysis for genomics batch correction. *Bioinformatics*. **30**:2757-63. PMID: 24907368. PMCID PMC4173013.
  3. Therapeutic resistance and precision medicine in cancer: I am PI on projects developing integrated computational-experimental strategies to infer the molecular mechanisms distinguishing therapeutic sensitivity from resistance from genomics data. Computationally, we inferred gene signatures related to therapeutic response all these projects with the CoGAPS method that we developed (2 below). We established the efficacy of this approach for precision medicine by demonstrating its accurate delineation specific cell signaling pathways from gene expression data in a controlled experimental model with activation of specific pathways in the EGFR network (a, featured in a commentary in *BMC Medicine*). This approach also inferred novel feedback mechanisms in which cells sensitive to anti-EGFR therapy have compensatory activation of growth factor receptor genes (b). Most recently, I formed a wet lab to test the role of dynamics in precision medicine. Specifically, we developed an innovative experimental protocol to trace the molecular state of HNSCC cancer cells over time as they acquire resistance to the anti-EGFR monoclonal antibody cetuximab. Coupling this technique to our novel data integration techniques enabled us to distinguish the relative timing of molecular changes, leading to the unanticipated finding that DNA methylation stabilizes but does not initiate acquired therapeutic resistance (c). A novel projection method developed to associate inferred gene signatures from one high-throughput dataset with data from new experimental conditions in (a) demonstrated that the gene signatures from our time course model distinguishes cetuximab response in HNSCC patients from gene expression data of their pre-treatment tumors. We have also applied similar methodology to other tumor types to demonstrate its generalizability to targeted therapeutic response. For example, when applied to RPPA data from the HPN-DREAM breast cancer network inference challenge, CoGAPS inferred that ligand stimulation had a greater impact on phospho-proteomic protein trajectories. This observation motivated our prediction algorithm for novel targeted agents, enabling me to lead the winning team in this DREAM8 contest subchallenge 2A (d).
    - a. **Fertig EJ**, Ren Q, Cheng H, Hatakeyama H, Dicker AP, Rodeck U, Considine M, Ochs MF, and Chung CH (2012) Gene expression signatures modulated by epidermal growth factor receptor activation and their relationship to cetuximab resistance in head and neck squamous cell carcinoma. *BMC Genomics*. **13**:160. PMID: 22549044. PMCID PMC3460736.
    - b. **Fertig EJ**, Ozawa H, Thakar M, Howard JD, Kagohara LT, Kringsfeld G, Ranaweera RS, Hughes RM, Perez J, Jones S, Favorov AV, Carey J, Stein-O'Brien G, Gaykalova DA, Ochs MF, Chung CH. (2016) CoGAPS matrix factorization algorithm identifies transcriptional changes in AP-2alpha target genes in feedback from therapeutic inhibition of the EGFR network. *Oncotarget*. **7**:73845-73864. PMID: 27650546. PMCID PMC5342018.
    - c. Stein-O'Brien G, Kagohara LT, Li S, Thakar M, Ranaweera R, Ozawa H, Cheng H, Considine M, Schmitz S, Favorov AV, Danilova LV, Califano JA, Izumchenko E, Gaykalova DA, Chung CH, **Fertig EJ**. (2018) Integrated time course omics analysis distinguishes immediate therapeutic response from acquired resistance. *Genome Research*. In press, preprint <https://doi.org/10.1101/136564>.
    - d. The HPN-DREAM Consortium. (2016) Inferring causal molecular networks: empirical assessment through a community-based effort. *Nature Methods*. **13**:310-318. PMID: 26901648. PMCID PMC4854847.
  4. Cancer genomics: I have led genomics analysis of datasets from numerous screening studies from diverse tumor types, experimental conditions, and high-throughput data platforms. In addition to analyses performed in my lab, I have devised a training plan to empower biologists in my collaborators' labs to perform their own bioinformatics analyses leading to numerous co-authored publications (e.g., a-c) and

their subsequent independent publications beyond the mentorship period. The majority of cancer genomics studies performed in my lab has focused on the epigenetic and transcriptional landscape of head and neck squamous cell carcinoma (HNSCC). This expertise is reflected in the first papers to characterize the landscape of gene fusions (a) and splice variation (b) in the HPV-positive HNSCC types. These studies found novel transcriptional alterations that compensate for pathway activity, not accounted for by mutations. Beyond these contributions to HNSCC genomics, I have also led similar analyses in other oral malignancies, lung, and breast cancers and applications to tumor immunology (c). These studies have extended beyond the genomics characterization of primary tumors, for example performing analyses of gene co-regulation of secreted factors by the tumor microenvironment in breast cancer (d). Together, these collaborative analyses demonstrate my expertise in inferring biologically-relevant findings from both proprietary and public domain high-throughput datasets to advance collaborative cancer research.

- a. Guo T, Gaykalova DA, Considine M, Wheelan S, Pallavajjala A, Bishop JA, Westra WH, Ideker T, Koch WM, Khan Z, **Fertig EJ**, and Califano JA. (2016). Characterization of functionally active gene fusions in human papillomavirus related oropharyngeal squamous cell carcinoma. *International Journal of Cancer*. **139**:373-382. PMID: 26949921. PMCID *in progress*.
  - b. Guo T, Sakai A, Afsari B, Considine M, Danilova L, Favorov AV, Yegnasubramanian S, Kelley DZ, Flam E, Ha PK, Khan Z, Wheelan SJ, Gutkind JS, **Fertig EJ**, Gaykalova DA, and Califano J. (2017) A Novel Functional Splice Variant of *AKT3* Defined by Analysis of Alternative Splice Expression in HPV-Positive Oropharyngeal Cancers. *Cancer Research*. **77**:5248-5258. PMID: 28733453. PMCID *in progress*.
  - c. Shuptrine CW, Ajina R, **Fertig EJ**, Jablonski SA, Kim Lyerly H, Hartman ZC, and Weiner LM. (2017) An unbiased in vivo functional genomics screening approach in mice identifies novel tumor cell-based regulators of immune rejection. *Cancer Immunology and Immunotherapy*. **66**:1529-1544. PMID: 28770278. PMCID PMC5854209.
  - d. Lee E, **Fertig EJ**, Jin K, Sukumar S, Pandey NB, Popel AS. (2014) Breast cancer cells condition lymphatic endothelial cells within pre-metastatic niches to promote metastasis. *Nature Communications*. **5**:4715. PMID: 25178650. PMCID PMC4351998.
5. Mathematical models of cellular signaling networks and complex systems: My unique, dynamical systems perspective on cancer bioinformatics was informed by my previous experience in applied mathematics. In addition to working with molecular data, I have developed novel mathematical models to predict the state of signaling pathways based upon their network structure. For example, I led formulation of a new model of coupled oscillators and switches (a) which demonstrated that signaling network dynamics must be inferred from their context in the broader network instead of from their structure in isolated motifs (b). Yet, mathematical models alone are insufficient to forecast the states of complex, dynamical systems. In weather forecasting, it is well established that regularly integrating mathematical models and measurements of a dynamical system improves predictions of future states. I developed novel algorithms to improve weather forecasts by incorporating indirect satellite observations into a flow-dependent data assimilation scheme, the local ensemble transform Kalman filter (LETKF, c). These algorithms were fundamental in transforming LETKF from a research data assimilation system into a semi-operational system (d), considered for making weather forecasts by the weather centers in Japan, Brazil, and the US Navy. As a result of this experience, I believe that predicting the state of any complex dynamical system, including cancer, must account for both observations of its current state and its dynamics. For example, I am currently PI on U01 and R01 projects that develops algorithms for personalized treatment selection that model changes in cellular signaling as a cancer cells respond to their microenvironment and therapy.
- a. **Fertig EJ**, Danilova LV, Favorov AV, and Ochs MF. (2011) Hybrid modeling of cell signaling and transcriptional reprogramming and its application in *C. elegans* development. *Front Genetic*. **2**:77. PMID: 22303372. PMCID PMC3268630.
  - b. Francis M and **Fertig EJ** (2012) Quantifying the dynamics of coupled networks of switches and oscillators. *PLoS One*. **7**:e29497. PMID: 22242172. PMCID PMC3252330.
  - c. **Fertig EJ**, Harlim J, and Hunt BR (2007) A comparative study of 4D-VAR and a 4D ensemble Kalman filter: Perfect model simulations with Lorenz-96. *Tellus A*. **59**:96-100.
  - d. Aravéquia JA, Szunyogh I, **Fertig EJ**, Kalnay E, Kuhl D and Kostelich EJ (2011) Evaluation of a strategy for the assimilation of satellite radiance observations with the local ensemble transform Kalman filter. *Monthly Weather Review*. **139**:1932-1951.

## Complete List of Published Work in MyBibliography:

<http://www.ncbi.nlm.nih.gov/sites/myncbi/1Rq7yzk6fclAe/bibliography/43872075/public/?sort=date&direction=descending>

### D. Ongoing Research Support

R01CA177669 (NCI) Fertig (PI) 09/16/14 – 06/30/19  
Dynamical Models of Cetuximab Resistance Drivers in HNSCC Based with Serial Omics Data  
This central goal of this project project is to develop novel computational algorithms to infer the molecular mechanisms underlying cetuximab resistance from *in vitro* and *in vivo* model of cetuximab resistant HNSCC.  
Role: PI

P30CA006973 (NCI) Nelson/Cope (PI) 05/01/06 - 04/30/22  
Regional Oncology Research Center: Bioinformatics Core  
The bioinformatics shared resource guarantees the availability of comprehensive Bioinformatics expertise to Cancer Center members.  
Role: Co-Investigator

U01CA196390 (NCI) Demarzo (PI) 09/10/15 – 08/31/20  
Multidiciplinary Integrative Genomic Approach to Distinguish Lethal from Indolent Prostate Cancer in Men of Europea and African Ancestry  
We first propose to perform an integrated, multi-dimensional genomic, epigenomic and expression analysis to uncover novel molecular pathways that characterize indolent vs. aggressive prostate cancers  
Role: Co-Investigator

R01CA200859 (NCI) Marchionni (PI) 04/05/16 - 03/31/21  
Hardwiring Mechanism into Predicting Cancer Phenotypes by Computational  
The major goal of this project is to develop novel methods to stratify cancer patients and predict their clinical course using biological information about the mechanisms underlying their disease.  
Role: Co-Investigator

90074231 Izumchenko (PI) 10/01/17 - 09/30/19  
Bonnie J Addario Lung Cancer Foundation  
Comprehensive analysis of the Genetic Landscape During Progression of Non-Small Cell Lung Adenocarcinoma  
This project will identify evolving clonal molecular events with WES and transcriptional data from clinical samples of lung cancer progression responsible for progression from preneoplastic lesions to cancer.  
Role: Co-Investigator

U01CA221007 Fertig (PI) 04/01/18 – 03/31/23  
NCI  
Integrating bioinformatics into multiscale models for hepatocellular carcinoma  
The major goal of this project is develop a novel hybrid, multiscale model that merges bioinformatics and mathematical models with organoid and mouse models for precision medicine in liver cancer.  
Role: PI

Fertig (PI) 04/01/18 – 03/31/19  
Silicon Valley Community Foundation  
Bayesian Sparse Matrix Factorization for Multimodal Integration  
This proposal will modify CoGAPS to a parallel framework for efficiency and to input the underlying estimates of transcript abundance, transcriptional variance, and sparsity characteristic of scRNA-seq data.  
Role: PI