## BIOGRAPHICAL SKETCH

Provide the following information for the Senior/key personnel and other significant contributors.
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 Professor of Oncology

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

| INSTITUTION AND LOCATION | DEGREE *(if applicable)* | Completion Date MM/YYYY | 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

The proposed project will develop an efficient bioinformatics algorithm (P-GAPS) to infer developmental trajectories and cell fate decisions in time course, multiomics bulk and single cell RNA-seq data. 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 distinct cancer subtypes and during acquired therapeutic resistance. My work on CoGAPS has been recognized through funding as a PI on K25 and R01 awards from NCI, the Johns Hopkins University Discovery Award, the Johns Hopkins University Catalyst Award and, as the best paper award at the 2012 IEEE BIBM Conference, and as leader of the winning team on the HPN-DREAM Breast Cancer Network Inference Challenge 2A; it was also featured in a commentary in BMC Medicine and an article published in *Time Magazine* and *US News and World Reports*. Yet, the CoGAPS MCMC algorithm is prohibitively computationally intensive for the extensive samples in the Human Cell Atlas and requires new methods to extend across data platforms. Implementation of an alternative parallel framework will improve its efficiency and platform-specific hyperparameters will enable its adaptation to single cell RNA-sequencing. To this work, I bring a strong foundation in applied mathematics for bioinformatics algorithm development and expertise in genomics analysis of human transcriptional data in complex study designs. Success in all these multi-disciplinary research activities has been a direct result of collaboration with scientists from diverse scientific backgrounds. As a result of these previous experiences, I am aware of the importance of frequent communication with members of the proposed collaborative network and HCA consortium to ensure well-integrated workflows and algorithm development. In summary, I have a demonstrated record of development of bioinformatics pattern detection algorithms supported through interdisciplinary collaboration. My expertise and experience have prepared me to develop the new parallel pattern detection algorithm, P-GAPS, to infer dynamic processes from multiomics, bulk and single cell data of retinal development.

### Positions and Honors

**Positions and Employment**

2004 – 2007 Research Assistantship, University of Maryland, College Park, MD,
2007 – 2008 Analyst, Metron, Inc., Reston, VA,
2008 – 2010 Research Fellow, Oncology Biostatistics, Johns Hopkins University School of Medicine, Baltimore, Maryland
2010 – 2013 Instructor, Oncology Biostatistics, Johns Hopkins University School of Medicine,

| | |
|---|---|
| | Baltimore, MD |
| 2011 – Present | Affiliate Faculty, Machine Learning, Johns Hopkins University, Baltimore, MD |
| 2011 – Present | Affiliate Faculty, Center for Computational Biology, Johns Hopkins University, Baltimore, MD |
| 2013 – Present | Assistant Professor, Oncology Biostatistics, The Johns Hopkins University School of Medicine, Baltimore, MD |
| 2015 – Present | Affiliate Faculty, Institute for Computational Medicine, Johns Hopkins University, Baltimore, MD |

**Honors**

| | |
|---|---|
| Oct 2012 | IEEE BIBM 2012 Best Paper Award |
| July 2013 | Awarded the Helen Masenhimer Fellowship through the Cleveland Foundation |
| Oct 2013 | Lead winning team on the HPN-DREAM Breast Cancer Network Inference Challenge 2A |
| 2013 – 2014 | Accepted participant in Emerging Women Leadership Program, Johns Hopkins University, Baltimore MD. |
| 2013 | Leader of winning team of DREAM8 HPN-DREAM Breast Cancer Network Inference Challenge, Subchallenge 2A. |
| 2014 | Winner of the Summer Science Program New Curriculum Prize. |
| 2016 | Awarded the Johns Hopkins University Discovery Award. |
| 2017 | Awarded the Johns Hopkins University IDIES Seed Grant |

**C. Contribution to Science**

1. Inferring meta-pathway activity from genomic data. I led development of a novel pattern identification algorithm, Coordinated Gene Activity in Pattern Sets (CoGAPS), 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, including principal component analysis and non-negative matrix factorization (d). As a result, I led the development of a novel statistic that used CoGAPS to refine databases of transcription factor targets (c), which was recognized as the best paper at the 2012 IEEE BIBM Conference. 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 phospho-proteomic trajectories with novel targeted agents, enabling me to lead the winning team in this DREAM8 contest subchallenge 2A (b). 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.

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: Epub ahead of print. PMID 28174896. PMCID in progress.
b. HPN-DREAM Consortium (2016) Inferring causal molecular networks: empirical assessment through a community-based effort. *Nature Methods*. Epub ahead of print. PMID: 26901648.
c. **Fertig EJ**, Favorov AV, and Ochs MF (2013) Identifying context-specific transcription factor targets from prior knowledge and gene expression data. *IEEE Trans Nanobioscience.* **12:**142-9. PMID: 23694699. PMCID PMC3759534.
d. Ochs MF and **Fertig EJ** (2012) Matrix factorization for transcriptional regulatory network inference. *IEEE Symp Comput Intell Bioinforma Comput Biol Proc.* **2012:**387-396. PMID: 25364782. PMCID PMC4212829.

2. Algorithms to standardize high throughput data. Robust data preprocessing techniques are essential to providing high throughput data of sufficient quality for subsequent analyses. I was PI on a study that developed a novel batch correction algorithm, permuted surrogate variable analysis (pSVA), to remove technical artifacts from high throughput data while preserving signal for pattern detection (a). This study also performed the first analysis of the effect of batch correction techniques on clustering and class prediction algorithms. We demonstrated that no single algorithm is optimal. Instead, batch correction algorithms must be tailored to the study to avoid either over or under-correcting high throughput data. I was also involved in developing a robust

algorithm to normalize DNA methylation data that accounts for the chemistry of Illumina 450K methylation arrays (b). This algorithm removed technical artifacts in DNA methylation data from head and neck samples in TCGA and has been implemented in a novel point and click, but reproducible web interface (c). I was also PI on a study to develop a reproducible, point and click web interface called AGA to automate analyses of DNA methylation and gene expression data (d).

a. Considine M, Parker H, Wei Y, Xia X, Cope L, Ochs MF, and **Fertig EJ** (2015) AGA: Interactive pipeline for reproducible gene expression and DNA methylation data analyses. *F1000Res.* **4**:28. PMID: 26535111. PMCID PMC4617321.
b. 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.
c. Fortin JP, Labbe A, Lemire M, Zanke BW, Hudson TJ, **Fertig EJ**, Greenwood CM, and Hansen KD (2014) Functional normalization of 450k methylation array data improves replication in large cancer studies. *Genome Biology.* **15**:503. PMID: 25599564. PMCID PMC4283580.
d. Fortin JP, **Fertig EJ**, and Hansen K (2014) shinyMethyl: interactive quality control of Illumina 450k DNA methylation arrays in R. *F1000Res*. **3**:175. PMID: 25285208. PMCID PMC4176427.

3. Cancer genomics from integrated genomics datasets. I am PI and lead bioinformatician on numerous projects to infer the molecular drivers and biomarkers that distinguish subtypes, treatment selection, and prognosis of head and neck cancers from high throughput data. I was PI on a study that developed a novel data integration algorithm that inferred epigenetic drivers in head and neck cancer subtypes to associate Hedgehog pathway activity with HPV-negative HNSCC tumors (a). I have also led analysis of data for biological models of HNSCC, which associated RAS pathway activity with acquired resistance to EGFR antibodies (b) and was featured in a commentary in BMC Medicine. In this role, I have also performed analyses from internal and public domain data that (c) characterized the mutational landscape of premalignant lesions relative to lung cancer and (d) led the analysis of TCGA data that confirmed the association of mutations with promoter methylation. Taken together, this body of work has suggested robust analyses from integrated genomics datasets in cancer and model systems.

a. Izumchenko E, Chang X, Brait M, **Fertig E,** Kagohara LT, Bedi A, Marchionni L, Agrawal N, Rajani R, Jones S, Hoque MO, Westra WH, and Sidransky D. (2015) Targeted sequencing reveals clonal genetic changes in the progression of early lung neoplasms and paired circulating DNA. Nat Commun. 2015;6:8528. PMID: 26374070, PMCID: PMC4595468.
b. Guerro-Preston R, Michailidi C, Marchionni L, Pickering CR, Frederick MJ, Myers JN, Yegnasubramanian S, Hadar T, Noordhuis MG, Zizkova V, **Fertig EJ**, Agrawal N, Westra W, Koch W, Califano JA, Velculescu VE, and Sidransky D. (2014) Key tumor suppressor genes inactivated by "greater promoter" methylation and somatic mutations in head and neck cancer. *Epigenetics*. **9**:1031-46. PMID: 24786473. PMCID PMC4143405.
c. **Fertig EJ**, Markovic A, Danilova LV, Gaykalova DA, Cope L, Chung CH, Ochs MF, and Califano JA (2013) Preferential activation of the hedgehog pathway by epigenetic modulations in HPV negative HNSCC identified with meta-pathway analysis. *PLoS One*. **8**:e78127. PMID: 24223768. PMCID PMC3817178.
d. **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.

4. Mathematical models of cellular signaling networks. In addition to developing algorithms to analyze 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 (b) which demonstrated that signaling network dynamics must be inferred from their context in the broader network instead of from their structure in isolated motifs (a). I also was lead developer of a new computational framework to model stochasticity and feedback in cellular signaling networks (c). These models will be integral in informing the algorithms that I am developing as PI on a study to infer mechanisms of targeted therapeutic resistance from high throughput time course data.

a. Taylor D, **Fertig EJ**, and Restrepo JG (2013) Dynamics in hybrid complex systems of switches and oscillators. *Chaos*. **23**:033142. PMID: 24089978. PMCID PMC3795755.

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**, 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.

5. Improving forecasts of evolving systems with algorithms that integrate big data into mathematical models.
My unique, dynamical systems perspective on cancer bioinformatics was informed by my previous experience in numerical weather prediction. 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, b-d). These algorithms were fundamental in transforming LETKF from a research data assimilation system into a semi-operational system (a), 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 an R01 that develops algorithms for personalized treatment selection that model changes in cellular signaling as a tumor acquires resistance.
a. 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.
b. **Fertig EJ**, Baek S-J, Hunt BR, Ott E, Szunyogh I, Aravéquia JA, Kalnay E, Li H, Liu J (2009) Observation bias correction with an ensemble Kalman filter. *Tellus A.* **61**:210-226.
c. **Fertig EJ,** Hunt BR, Ott E, and Szunyogh I (2007) Assimilating non-local observations with a local ensemble Kalman filter. *Tellus A.* **59**:719-730.
d. **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.

**Complete List of Published Work in MyBibliography:**
**http://www.ncbi.nlm.nih.gov/sites/myncbi/1Rq7yzk6fclAe/bibliography/43872075/public/?sort=date&direction=descending**

## D. Research Support

### Ongoing Research Support

JHU Catalyst Award — Fertig (PI) — 07/01/17 – 06/31/18
Modeling intra-tumor heterogeneity during acquired therapeutic resistance
This project aims to adapt non-negative matrix factorization to scRNA-seq data from time course, single cell data in acquired therapeutic resistance in cancer.
Role: PI

JHU IDIES seed grant — Fertig (PI) — 03/01/17 – 02/28/18
Variational Bayes Gene Activity in Pattern Sets (VB-GAPS) Bioinformatics Algorithm for Efficient Precision Medicine in Oncology
This project aims to develop new variational techniques for efficient, smooth-sparse non-negative matrix factorization in genomics.
Role: PI

NCI R01CA200859 — 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

NCI U01CA196390 — Schaeffer (PI) — 09/10/15 – 08/31/20
Multidiciplinary Integrative Genomic Approach to Distinguish Lethal from Indolent Prostate Cancer in Men of Europena and African Ancestry

Role: Co-Investigator

NIDCR P50DE019032 Fertig/Sidransky (PI) 08/01/15 – 07/31/17
SPORE in Head & Neck Cancer: PILOT
Expression Variability Analysis for Alternative Splice Events in HNSCC
The major goal of this SPORE project is thus to reduce the morbidity and mortality of head and neck cancer through a highly coordinated program consisting of basic and clinical research focusing on the molecular parameters associated with the development and progression of this deadly disease.
Role: Co-Investigator

NIDCR R21DE025398 Gaykalova (PI) 07/01/15 – 06/30/17
Characterizing genome-wide alternative splicing in HPV related HNSCC
This project will provide insight into HPV+ HNSCC biology and will potentially provide therapeutic and prognostic targets for the development of HPV+ HNSCC specific therapies
Role: Co-Investigator

NCI R01CA177669 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

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

**Completed Research Support**

Johns Hopkins University Fertig (PI) 07/01/16 – 06/31/16
Discovery Award
Systems biology of invasive liver cancer
Role: PI

Johns Hopkins SPORE Fertig (PI) 01/01/17 – 07/31/18
SPORE in Head & Neck Cancer: PILOT
Gene expression and chromatin structure changes in HNSCC and their role in acquired resistance to cetuximab.
Role: PI

Johns Hopkins University SPORE Fertig (PI) 09/01/10 – 08/31/12
SPORE in Head & Neck Cancer: PILOT
Inferring MAPK signaling in EGFR sensitive and resistant head and neck cancers
Role: PI

NIDCR P50DE019032 Fertig (PI) 08/01/15 – 07/31/16
SPORE in Head & Neck Cancer: PILOT
Automated Genomics Analysis for Head and Neck Cancer
Role: Co-Investigator

NCI K25CA141053 Fertig (PI) 07/01/10 – 06/30/15
Identifying Malignant Cell Signaling from Protein Interactions an Polyomic Data
Role: PI

**Description of other key personnel**

**Alexander V Favorov, Research Associate** Dr. Favorov is specialized in Markov chain Monte Carlo (MCMC) techniques for genomics. He has developed several state of the art algorithms for motif discovery and genome-wide correlations. He has a long history of collaboration with Dr. Fertig on CoGAPS. He will collaborate closely with postdoctoral fellow Genevieve Stein-O'Brien to develop the distributions to model the sparsity hyperparameter in each bulk and single cell genomics technology (Aim 2).

**Thomas Sherman, Biostatistician** Mr. Sherman completed his MS in Applied Mathematics and Statistics at Johns Hopkins University. He works as a programmer in Dr. Fertig's lab, maintaining CoGAPS and implementing efficient data structures. He will be responsible for implementing the message passing required for the parallelization methods proposed in Aim 1.

**Genevieve Stein-O'Brien, Postdoctoral Fellow** Dr. Stein-O'Brien completed her PhD from the Institute for Human Genetics in 2017. For her dissertation work, she applied CoGAPS to analyze numerous developmental datasets. She has developed new parallelization methods and gene selection techniques for genome-wide CoGAPS. Currently, Dr. Stein-O'Brien is a postdoctoral fellow co-supervised by Dr. Fertig and collaborative network member Dr. Goff. She is listed in both proposals. In this proposal, Dr. Stein-O'Brien is listed as responsible for the algorithm development and analyses proposed in this award. Her work will be completed in collaboration with all key personnel on this proposal and co-supervised by Dr. Fertig and Dr. Goff. If both Dr. Fertig and Dr. Goff's awards are funded, a TBD postdoc will be hired to collaborate with Dr. Stein-O'Brien on these efforts.