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: Stuart, Joshua M
eRA COMMONS USER NAME (credential, e.g., agency login): JOSHSTUART
POSITION TITLE: Professor
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
University of Colorado, Boulder, CO	B.A.	12/96	Molecular Biology
University of Colorado, Boulder, CO	B.S.	12/96	Computer Science
University of Colorado, Boulder, CO	R.A.	08/98	Chaos Theory
Stanford University, Stanford, CA	Ph.D.	01/2004	Biomedical Informatics

A. Personal Statement

Dr. Stuart has expertise in developing computational models to integrate multiple sources of information and a background in machine-learning applied to high-throughput datasets. He has recently developed pathway-based models to integrate multiple sources of gene activity to predict alterations and clinical outcomes in tumor samples. He co-leads the Pan-Cancer working groups for the Cancer Genome Atlas project and the International Cancer Genomics Consortium, co-directs the UCSC-Buck institute genome data analysis center, directs a bigdata resource for NCI to develop high-level information from CGHub raw DNA/RNA sequence data, and leads the pathway analysis for a prostate cancer Stand Up To Cancer and Prostate Cancer Foundation “Dream Team.”

B. Positions and Honors

Positions and Employment

1993-1996 Laboratory Research Assistant, Dept. of Molecular Biology (Dr. G. Stormo), University of Colorado, Boulder, Colorado.
1994 University of Colorado Health Science Cancer Fellowship, Boulder Colorado, Summer.
1996-1997 Research Assistant, Dept. of Computer Science (Dr. L. Bradley), UC, Boulder.
2000 Teaching Assistant in Biomedical Informatics, Stanford, University, Stanford, CA.
2003-2009 Assistant Professor, Dept. of Biomolecular Engineering, University of California, Santa Cruz.
2009-2013 Associate Professor, Dept. of Biomolecular Engineering, University of California, Santa Cruz.
2013-*present* Professor, Dept. of Biomolecular Engineering, University of California, Santa Cruz.

Honors

1995-1996 Achievement Rewards for College Scientists (ARCS) scholarship recipient for research in Dr. G. Stormo’s laboratory.
1996 magna cum laude, MCD Biology, University of Colorado.
2006 University of Colorado Kalpana Chawla Outstanding Recent Graduate Award.
2006-2014 Alfred P. Sloan research fellowship.
2009-2014 NSF CAREER Award.
2013-present Jack Baskin Endowed Chair, UCSC School of Engineering

C. Contribution to Science

My UCSC lab has developed leading probabilistic pathway integration methods to infer genetic networks and gene activities from many types of data collected on cancer samples (Vaske, *PloS Comp Bio* 2009, Vaske Bioinformatics 2010, Ng, Bioinformatics 2012). Of note is our key contribution, in collaboration with David Haussler, in the development of a rich mathematical model called PARADIGM that can incorporate virtually any kind and number of genomics, functional-genomics, or epigenomics data together in a unified analysis.

The methods are widely used by the national TCGA project, international ICGC project, and are now being deployed in multiple translational applications including several Stand Up To Cancer Dream team projects. In the past several years, the PARADIGM algorithm has contributed key results to over a dozen of the major cancer analyses published or to appear in *Nature* and *Cell*, primarily by the TCGA project. In addition, we have developed a method called TieDIE (Paull, *Bioinformatics*, 2013) to link genomic events with altered gene expression pathways.

1. Vaske CJ, House C, Luu T, Frank B, Yeang CH, Lee NH, Stuart JM. "A Factor Graph Nested Effects Model to Identify Networks from Genetic Perturbations." Public Library of Science Computational Biology. 2009. Jan 5. e1000274. PMCID: PMC2613752.
2. C. J. Vaske, Benz, S. C., Sanborn, Z. J., Earl, D., Szeto, C., Zhu, J., Haussler, D., and Stuart, J. M., "Inference of patient-specific pathway activities from multi-dimensional cancer genomics data using PARADIGM", Bioinformatics, vol. 26, pp. i237-i245, 2010.
3. S. Ng, Collisson, E. A., Sokolov, A., Goldstein, T., Gonzalez-Perez, A., Lopez-Bigas, N., Benz, C., Haussler, D., and Stuart, J. M., "PARADIGM-SHIFT predicts the function of mutations in multiple cancers using pathway impact analysis", Bioinformatics, vol. 28, pp. i640-i646, 2012.
4. Paull EO, Carlin DE, Niepel M, Sorger PK, Haussler D, and Stuart JM. "Discovering causal pathways linking genomic events to transcriptional states using Tied Diffusion Through Interacting Events (TieDIE)." Bioinformatics (2013) doi: 10.1093/bioinformatics/btt471, Aug. 2013.

As Pan-Cancer project lead for the TCGA project, my role is to create a unified vision for the project, breakdown the project into sub-goals with assigned leads, coordinate all data, and schedule all downstream analyses. The first Pan-Cancer project (2012-2013) detailed the analysis of over 5000 patient samples covering twelve different forms of cancer. I oversaw the publication of 28 manuscripts, and wrote the perspective paper (Weinstein, *Nat Genetics* 2013) that described the overall project. I also led a group to collect all of the datasets produced by the six different data platforms in order to identify patient subtypes suggested by each of the data modalities (Hoadley, *Cell* 2014). The surprising finding was that all platforms were driven strongly by the tissue of origin but that 10% of the tumors were connected to different tissues than from which they originated. I now co-lead a new Pan-Cancer project for the International Cancer Genome Consortium to identify connections across tumor types focused on the analysis of the non-coding parts of the genome.

1. The Cancer Genome Atlas Research Network, Weinstein JN, Collisson EA, Mills GB, Shaw KRM, Ozenberger BA, Ellrott K, Shmulevich I, Sander C, **Stuart JM**. "The Cancer Genome Atlas Pan-Cancer analysis project." *Nature Genetics* 45(10):1113-20, 9/2013.
2. Assessing the clinical utility of cancer genomic and proteomic data across tumor types. Yuan Y, et al. *Nat Biotechnol*. 2014 Jul;32(7):644-52. doi: 10.1038/nbt.2940. Epub 2014 Jun 22. PMID: 24952901.
3. The International Cancer Genome Consortium, "International network of cancer genome projects", *Nature*, vol. 464, pp. 993 - 998, 2010.
4. Hoadley KA, et al."Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin." *Cell*. 2014 Aug 14;158(4):929-44. doi: 10.1016/j.cell.2014.06.049. Epub 2014 Aug 7. PMID: 25109877.

My lab develops new machine-learning methods to predict cancer outcomes more accurately. Recently, we have developed a new method called GELNets that uses gene-gene pathway information to extend classic statistical regression (Sokolov, *PLoS Comp Bio* 2016). We worked with the Witte Lab (HHMI, UCLA) to apply our work to show that stem cell signatures characterize certain metastatic prostate cancers (Smith, *PNAS*, 2015). We have developed a new way to develop biomarkers that allows them to recognize small quantities of a subclone in a complex tumor using so-called "one-class" methods (Sokolov, *Proceedings of Pacific Symposium on Biocomputing 2016*). We have also developed approaches to predict what genes are *essential* in cancer cells.

We compete in open competitions run by DREAM. Most recently, we won the Broad Institute's Achilles challenge to predict genes essentiality across a diverse set of cell lines (Hill, *Nat Methods*, 2016). These methods are critical for modeling cancer cells and for identifying their vulnerabilities that we can then target for

treatment.. Our method makes use of pathway- and gene-module-based information to develop a multiple kernel learning approach that is more accurate than hundreds of other approaches submitted to this challenge. We are extending our efforts to incorporate additional types of data into the assessment of patient outcomes that augment the genomics.

1. Pathway-Based Genomics Prediction using Generalized Elastic Net. Sokolov A, Carlin DE, Paull EO, Baertsch R, Stuart JM. *PLoS Comput Biol*. 2016 Mar 9;12(3). PMID: 26960204.
2. One-class detection of Cell States in Tumor Subtypes. Sokolov A, Paull EO, **Stuart JM**. *Pac Symp Biocomput.* 2016;21:405-16. PMID: 26776204.
3. A basal stem cell signature identifies aggressive prostate cancer phenotypes. Smith BA, Sokolov A, Uzunangelov V, Baertsch R, Newton Y, Graim K, Mathis C, Cheng D, **Stuart JM**, Witte ON. *Proc Natl Acad Sci* U S A. 2015 Nov 24;112(47):E6544-52. PMID: 26460041.
4. Inferring causal molecular networks: empirical assessment through a community-based effort. Hill SM, DREAM Consortium. *Nat Methods*. 2016. Apr;13(4):310-318. PMID: 26901648

My lab works toward developing new ways to allow biologists to navigate the results of high-throughput assays that build on the topomap displays and search engines mentioned above. I have worked on network display views in collaboration with Charles Delisi's group (Hu et al *Nat Methods 2007*), and Bruce Conklin's group on GenMapp (that later became WikiPathways). We have also developed CircleMap displays to display coordinated views of many genes in their network context with all associated data (Wong, *NAR* 2013). With David Haussler, I have collaborated to develop the UCSC Cancer Genomics Browser (Sanborn et al. *NAR* 2011).

1. Hu Z, Mellor J, Wu J, Kanehisa M, **Stuart JM**, Delisi C. "Towards zoomable multidimensional maps of the cell." *Nat Biotechnol*. 2007 May;25(5):547-54. No PMCID.
2. Salomonis N, et al. "GenMAPP 2: New Features and Resources for Pathway Analysis." *BMC Bioinformatics* 8:217. June (2007).
3. Z. J. Sanborn, Benz, S. C., Craft, B., Szeto, C., Kober, K. M., Meyer, L., Vaske, C. J., Goldman, M., Smith, K. E., Kuhn, R. M., Karolchik, D., Kent, J. W., **Stuart, J. M**., Haussler, D., and Zhu, J., "The UCSC cancer genomics browser: update 2011", *Nucleic Acids Research*, vol. 39, pp. D951-D959, 2011.
4. Wong CK, Vaske CJ, Ng S, Sanborn JZ, Benz SC, Haussler D, **Stuart JM**. "The UCSC Interaction Browser: multidimensional data views in pathway context." *Nucleic Acids Res*. 2013.

My lab is active in organizing and competing in open competitions to find the best algorithms for important problems facing the biomedical community. These competitions post training datasets and evaluate submitted predictions using private sets of validation data to which only the organizers have access. I have partnered with DREAM, the leading organizing body for these open competitions, to launch several cancer-genomics-related challenges around the theme of interpreting high-throughput DNA and RNA sequence reads. Our first challenge, the DREAM ICGC/TCGA Somatic Mutation Calling challenge was a major success with now over 400 contestants and over 1000 submitted predictions (Boutros *et al Nature Genet*. 2014.. Surprisingly, we found a common set of false positives made by many of the top methods that matches a well-known mutation signature (Ewing *et al Nat Biotech* 2015).

1. Boutros PC, Ewing AD, Ellrott K, Norman TC, Dang KK, Hu Y, Kellen MR, Suver C, Bare JC, Stein LD, Spellman PT, Stolovitzky G, Friend SH, Margolin AA, **Stuart JM**. "Global optimization of somatic variant identification in cancer genomes with a global community challenge". *Nat Genet*. 2014 Apr;46(4):318-9. doi: 10.1038/ng.2932. PMID: 24675517; PMCID: PMC4035501.
2. Boutros PC, Margolin AA, **Stuart JM**, Califano A, Stolovitzky G. "Toward better benchmarking: challenge-based methods assessment in cancer genomics." *Genome Biol*. 2014 Sep 17;15(9):462. PMID: 25314947
3. Ewing AD, et al. "Combining accurate tumour genome simulation with crowd-sourcing to benchmark somatic single nucleotide variant detection." *Nat Biotech.* 2015. Jul;12(7):623-30. PMID: 25984700. PMCID: PMC4035501.

List of Biomedical-related Published Works in MyBibliography:	List of all Published Works in Google Scholar:
http://www.ncbi.nlm.nih.gov/myncbi/collections/bibliography/41171914	https://scholar.google.com/citations?user=-corgUwAAAAJ&hl=en

D. Research Support

Ongoing Research Support

5U24CA143858-05 REVISED (Haussler, Stuart) 9/28/09 – 6/30/16 NCE
NIH/NCI
UCSC-Buck Institute Genome Data Analysis Center for the Cancer Genome Atlas Research Network
Develop and apply methods for high throughput production-ready processing and analysis of large-scale next-generation sequencing data produced by the CGA project.

SU2C-AACR-DT0812, subaward 7467sc-05 (Small, Stuart) 1/1/2013-12/31/2016 NCE
Prostate Cancer Foundation/AACR/SU2C/UCSF
Targeting Adaptive Pathways in Resistant CRPC
Project supports identifying pathways in prostate cancer underlying androgen inhibition resistant disease. Dr. Stuart's lab will develop novel algorithms and deploy a data structure called to link together findings across labs.

5R01CA180778-03 (Stuart) 6/1/2013 – 5/31/2018
National Institutes of Health/NCI
BigData: Mid-Scale:DCM: DA: ESCE: Discovering Molecular Processes and Patient Outcome Patterns in Large-Scale Cancer Genomics Datasets Using a Biomedical Evidence Graph
Create a resource of interpretive levels of data derived from next generation sequencing data deposited in the Cancer Genomics Hub (CGHub).

GC1R-06673B, subaward: PO145602 (Ideker, Stuart) 6/1/14-5/31/16
Salk/CIRM
CIRM Center for Genomic Excellence Center Initiated Project
Develop a suite of bioinformatics tools and resources for advanced analysis of 'omics data generated by the CIRM Genome Center, with the goals of formulating molecular network models and for guiding predictions of cell fate. The primary tools, which are general, will be developed and applied together with investigators from the other center-initiated and collaborative projects.

GC1R-06673C (Stuart, Haussler) 6/1/14-5/31/19
CIRM
CIRM Center for Genomic Excellence
Create the "Stem Cell Hub (SCHub)" to manage next-generation sequencing datasets for the California Institute of Regenerative Medicine.

5R01GM109031-02 (Stuart, Haussler) 9/15/14-6/30/19
NIH
New Integrative Pathway Analysis Methods to Predict Biomedical Outcomes
Extend machine-learning and probabilistic graphical modeling approaches to reveal common mechanisms of stem cells and tumor biology to shed light on new treatment options for cancer.

5U54HL127365-02, subaward: 152296.5089286.0207 (Sorger, Stuart) 9/10/14-6/30/20
NIH/NHLBI/Harvard
Pharmacological Response Signatures and Disease Mechanism
Development of computational algorithms and software to interface the UCSC Cancer Genomics Browser tools with the LINCS project. Work will include expansion of the GELNets machine-learning framework to identify pathway-driven predictors of drug and ligand-induced responses in the Sorger cell line studies, and adaptation of the TieDIE algorithm to work with the RPLA data collected in the Sorger lab.