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.	12/1996	Molecular Biology
University of Colorado, Boulder, CO	B.S.	12/1996	Computer Science
University of Colorado, Boulder, CO	R.A.	08/1998	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, leads the pathway analysis for a prostate cancer Stand Up To Cancer and Prostate Cancer Foundation “Dream Team,” and directs the development of single cell data analysis portals for the California Institute for Regenerative Medicine’s Genome Center of Excellence and the Chan-Zuckerberg Institute’s Human Cell Atlas.

Training. I have been mentoring students and postdocs for sixteen years, and the success of my trainees is paramount to me. The 15 PhD students I have trained have ended up in top-notch postdoctoral labs -- Princeton, the Whitehead Institute, UC San Diego, University of Toronto, and Columbia to name a few -- or have started their own labs (Univ of Cincinnati) or taken research positions in industry -- Agilent, Illumina, and BMS. Two of my PhD students formed a startup company, now in its 8th year, to use genomics and systems biology technologies to find new treatments for cancer. One former postdoc now runs the biostatistics core at Harvard’s Systems Biology Center and the other founded a startup company in Silicon Valley to identify new drugs using computational approaches. I develop personalized training support for each individual trainee and strive to provide the appropriate level of mentoring. For example, not every student or postdoc requires the same level of training; cultivating a trainee’s independence while ensuring sufficient support is critical. In the publications below I have indicated authors that were former students or postdocs and bolded those that are now in positions in biotech. Many graduate students have played keyed roles, and earned significant authorship, on high-profile publications in collaborations or as part of consortium work. Most notably in the national pan-cancer studies investigating cancer exomes published by *Cell* in 2018 and most recently in the international analysis of whole genomes of cancer published in February 2020 by *Nature*. Graduate students also developed our highly used TumorMap (Newton et al., *JCO* 2018), and developed a network diffusion method to predict treatments in metastatic prostate cancer (Drake et al. *Cell* 2018).

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

1. The Stuart 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. Our method was most recently used to link phosphoproteomics data, transcriptional data, and mutations to create personalized networks for late stage prostate cancer samples providing clues into treatment options (Drake, *Cell*, 2016).

a) **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.
b) **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.
c) 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.
d) Drake, Justin M., Evan O. Paull, Nicholas A. Graham, John K. Lee, Bryan A. Smith, Bjoern Titz, Tanya Stoyanova, et al. 2016. "Phosphoproteome Integration Reveals Patient-Specific Networks in Prostate Cancer." *Cell* 166 (4): 1041–54.

2. As lead of the TCGA Pan-Cancer project, my role is to create a unified vision, coordinate all data, and delegate 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 to identify patient subtypes suggested by each of the data modalities (Hoadley, *Cell* 2018). The surprising finding was that 10% of the tumors grouped with other tumors that were from different tissues of the body. I have also co-lead the most recent 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. I led the analysis to investigate the pathways and networks connecting non-coding and coding alterations (Reyna, *Nature Comm*, 2020) as well as the marker study to investigate how tumors maintain immortality through telomere maintenance strategies (Campbell, *Nature*, 2020).

a) 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.

b) Hoadley KA Yau C Hinoue T Wolf DM Lazar AJ Drill E Shen R Taylor AM Cherniack AD Thorsson V Akbani R Bowlby Wong CK Wiznerowicz M Sanchez-Vega F Robertson AG Schneider BG Lawrence MS Noushmehr H Malta TM, The Cancer Genome Atlas Network, Stuart JM, Benz CC, Laird PW. "Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer" *Cell*. Volume 173 Issue 2 p291–304.e6 5 April 2018
c) Reyna, M.A., Haan, D., Paczkowska, M. et al. Pathway and network analysis of more than 2500 whole cancer genomes. *Nat Commun* 11, 729 (2020)
d) Campbell, P.J., Getz, G., Korbel, J.O. et al. Pan-cancer analysis of whole genomes. *Nature* 578, 82–93 (2020).

3. 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; Gonen, *Cell Syst,* 2017). 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.

a) Smith BA, Sokolov A, **Uzunangelov V**, Baertsch R, **Newton Y**, Graim K, Mathis C, Cheng D, Stuart JM, Witte ON. A basal stem cell signature identifies aggressive prostate cancer phenotypes. *Proc Natl Acad Sci* U S A. 2015 Nov 24;112(47):E6544-52. PMID: 26460041.
b) Hill SM, Heiser LM, Cokelaer T, Unger M, Nesser NK, Carlin DE, Zhang Y, Sokolov A, Paull EO, Wong CK, Graim K, Bivol A, Wang H, Zhu F, Afsari B, Danilova LV, Favorov AV, Lee WS, Taylor D, Hu CW, Long BL, Noren DP, Bisberg AJ; HPN-DREAM Consortium, Mills GB, Gray JW, Kellen M, Norman T, Friend S, Qutub AA, Fertig EJ, Guan Y, Song M, Stuart JM, Spellman PT, Koeppl H, Stolovitzky G, Saez-Rodriguez J, Mukherjee S.Inferring causal molecular networks: empirical assessment through a community-based effort. *Nat Methods*. 2016. Apr;13(4):310-318. PMID: 26901648.
c) Graim K, Liu TT, Achrol AS, Paull EO, **Newton Y**, Chang SD, Harsh GR 4th, **Cordero SP**, Rubin DL, Stuart JM. Revealing cancer subtypes with higher-order correlations applied to imaging and omics data. *BMC Med. Genomics* 10 20 (2017). PMID: 28359308 PMCID: PMC5374737.
d) Gönen M, Weir BA, Cowley GS, Vazquez F, Guan Y, Jaiswal A, Karasuyama M, **Uzunangelov V**, Wang T, Tsherniak A, Howell S, Marbach D, Hoff B, Norman TC, Airola A, Bivol A, Bunte K, Carlin D, Chopra S, Deran A, Ellrott K, Gopalacharyulu P, Graim K, Kaski S, Khan SA, **Newton Y**, **Ng S**, Pahikkala T, Paull E, Sokolov A, Tang H, Tang J, Wennerberg K, Xie Y, Zhan X, Zhu F. A Community Challenge for Inferring Genetic Predictors of Gene Essentialities through Analysis of a Functional Screen of Cancer Cell Lines. *Cell Syst*. 2017 Oct 4. pii: S2405-4712(17)30392-7. doi: 10.1016/j.cels.2017.09.004. PMID: 28988802 PMCID: PMC5814247

4. 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. We now have a focus on interpreting single cell RNA-sequencing experiments and published a paper to identify "cell trajectories" with a novel algorithm (Cordero, *PSB* 2017). 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). Recently, we developed the UCSC TumorMap to display cancer samples in a familiar GoogleMaps display (Newton et al. *Cancer Res.* 2017).

a) **Newton Y**, Novak AM, Swatloski T, et al. TumorMap: Exploring the Molecular Similarities of Cancer Samples in an Interactive Portal. *Cancer Res*. 2017;77(21):e111-e114. doi:10.1158/0008-5472.CAN-17-0580
b) **Cordero P**, Stuart JM. Tracing co-regulatory network dynamics in noisy, single-cell transcriptome trajectories. *Proceedings of the Pacific Symposium on Biocomputing*. January 2017.PMID: 27897008, PMCID: PMC5203771
c) 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.

d) 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.

5. 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).

a) 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.
b) 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
c) 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.

D. Research Support

Active

NIH/NCI 5U24CA210990-02 (Stuart) 9/15/16-8/31/21
UCSC-Buck Specialized Genomic Data Center for the Genomic Data Analysis Network
The project will implement and sustain the UCSC-Buck Specialized Genomic Data Center for the Genomic Data Analysis Network

NIH/NIGMS 5R01GM109031-05 (Stuart, Haussler) 9/15/14-6/30/20 (NCE)
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.

00093733.0 (Stuart) 4/1/19-3/31/22
Seagate Technology LLC
Computational Storage for Accelerated HCA Data Repository
To explore and deploy a much more sophisticated infrastructure using the Human Cell Atlas gene-cell matrix dataset as a test bed for compute capable storage and provide a fast mechanism to extract gene expression signatures associated with any particular cluster.

Recently completed

CIRM GC1R-06673-C (Stuart, Haussler) 6/1/14-5/31/20 (NCE)
Center of Excellence for Stem Cell Genomics - UCSC
Create the "Stem Cell Hub (SCHub)" to manage next-generation sequencing datasets for the California Institute of Regenerative Medicine.

NIH/NCI 5R01CA180778-05 (Stuart, Haussler) 6/1/13-5/31/19 (NCE)
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).

Chan Zuckerberg Initiative/SVCF 2018-182800 (Stuart) 3/1/18-2/28/28 (NCE)
Unified classification of cellular processes through systematic, cross-species aggregation of single-cell gene expression trajectories

Developing standards for describing “cell trajectories” found in single cell RNA-Seq datasets, methods to align trajectories, and tools to search known and novel cell differentiation transitions.

CIRM/Salk GC1R-06673-B, subaward: PO145602 Mod03 (Ideker, Stuart) 6/1/14-5/31/20 (NCE)
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.

DoD/UCSF W81XWH-16-1-0495 (Small, Stuart) 9/1/16-8/31/19
Identifying Targetable Adaptive Pathways in Abiraterone and Enzalutamide Refractory Intermediate Atypical Carcinoma
The goal of this proposal is to both identify the activated adaptive pathways present in Intermediate Atypical Cancer (IAC) form of metastatic Castration Resistant Prostate Cancer (mCRPC), and to delineate the relative activity of the androgen receptor (AR) pathway. Ultimately, we aim to develop co-targeting clinical strategies inhibiting AR and adaptive pathway as a means of delaying and/or arresting the transdifferentiation from adenocarcinoma to IAC/SCNC (small cell/neuroendocrine prostate cancer). UCSC will carry out the development and application of integrated pathway analysis for this project.