Sean Whalen, Ph.D.

Gladstone Institutes
1650 Owens Street
San Francisco, CA 94158

Email: shwhalen at gmail.com
Personal: node99.org
Professional: linkedin.com/in/shwhalen

Current Position

2025-   Principal Scientist: AI & Genomics, Gladstone Institutes

Research Interests

Machine learning in gene regulation, functional genomics, evolution, and 3D genome organization

Programming Languages

Python (pytorch, keras, scikit-learn, pandas), R (caret, tidyverse, bioconductor), Matlab, Java, C, C++, CUDA, OpenGL, HTML/CSS/JavaScript

Appointments

2017-2024   Senior Research Scientist, Gladstone Institutes
2013-2016   Biostatistician, Gladstone Institutes
2012-2013   Postdoctoral Fellow, Institute for Genomics and Multiscale Biology, Mount Sinai School of Medicine, Advisor: Gaurav Pandey
2011-2012   Postdoctoral Researcher, DARPA Mission-oriented Resilient Clouds, Intrusion Detection Systems Lab, Columbia University, Advisor: Salvatore J. Stolfo
2010-2011   I3P Postdoctoral Fellow, Computational Research Division, Lawrence Berkeley Natl. Lab
2010   Researcher, Anomaly Detection in High Performance Computing, UCDavis Security Lab
2009   Teaching Assistant, Operating Systems, Instructor: Daniela Oliviera
2008-2009   Researcher, Immersive Network Visualization, William M. Keck Foundation
2007   Teaching Assistant, Artificial Intelligence, Instructor: Ian Davidson
2006   Teaching Assistant, Operating Systems, Instructor: Felix Wu
2006   Teaching Assistant, Introduction to Object Oriented Programming, Instructor: Sean Davis
2004-2006   Researcher, Vulnerability Analysis, UCDavis Security Lab

Education

2010   Ph.D. in Computer Science, University of California, Davis
            Advisors: Matt Bishop (Computer Security) and James P. Crutchfield (Physics)
2006   M.S. in Computer Science, University of California, Davis
2002   B.S. in Computer Science (Cum Laude), University of Nebraska, Omaha

Grants, Honors & Awards

2024   Gladstone Excellence in Science Award
2013   Center for Technology, Innovation, and Entrepreneurship Grant Recipient
2010-2011   Institute for Information Infrastructure Protection Postdoctoral Fellowship ($150k)
2003   Graduate Assistance in Areas of National Need Doctoral Fellowship (Tuition)
1998-2002   Walter Scott Jr. Scholarship, University of Nebraska Foundation (Tuition, Board)

Peer-Reviewed Publications

Papers

2025   "Cis-Regulatory Element and Transcription Factor Circuitry Required for Cell-Type Specific Expression of FOXP3," Jennifer Umhoefer, Maya Arce, **Sean Whalen**, et al., submitted.

2024   "Deep learning prediction of gene expression from personal genomes," Shiron Drusinsky, **Sean Whalen** (co-first), Katherine S. Pollard, submitted.

2024   "Massively parallel characterization of psychiatric disorder-associated and cell-type-specific regulatory elements in the developing human cortex, Chengyu Deng, **Sean Whalen** (co-first), Marilyn Steyert et al., *Science*.

2023   "Three-dimensional genome re-wiring in loci with Human Accelerated Regions," Kathleen C. Keough, **Sean Whalen**, Fumitaka Inoue et al., *Science*.

2023   "Machine learning dissection of Human Accelerated Regions in primate neurodevelopment," **Sean Whalen** (co-first), Fumitaka Inoue, Hane Ryu et al., *Neuron*.

2023   "An atlas of lamina-associated chromatin across twelve human cell types reveals an intermediate chromatin subtype," Parisha P. Shaw et al., *Genome Biology*.

2022   "Single Cell Epigenetics Reveal Cell-Cell Communication Networks in Normal and Abnormal Cardiac Morphogenesis," Sanjeev S. Ranade, **Sean Whalen**, Ivana Zlatanova et al., *in review*.

2022   "Enhancer Function and Evolutionary Roles of Human Accelerated Regions, **Sean Whalen** and Katherine S. Pollard, *Annual Reviews of Genetics*.

2021   "Navigating the pitfalls of applying machine learning in genomics," **Sean Whalen** (co-first), Jacob Schreiber, William S. Noble, and Katherine S. Pollard, *Nature Reviews Genetics*.

2021   "Autism risk gene POGZ promotes chromatin accessibility and expression of clustered synaptic genes," Eirene Markenscoff-Papadimitriou, Fadya Binyameen, **Sean Whalen** et al., *Cell Reports*.

2020   "A Chromatin Accessibility Atlas of the Developing Human Telencephalon," Eirene Markenscoff-Papadimitriou, **Sean Whalen** (co-first), Pawel Przytycki et al., *Cell*.

2020   "lentiMPRA and MPRAflow for high-throughput functional characterization of gene reg-

ulatory elements," M. Grace Gordon et al., *Nature Protocols.*

2019   "AlleleAnalyzer: a tool for personalized and allele-specific sgRNA design," Kathleen C. Keough et al., *Genome Biology.*

2019   "The Glycan CA19-9 Promotes Pancreatitis and Pancreatic Cancer," Dannielle Engle et al., *Science.*

2019   "Most chromatin interactions are not in linkage disequilibrium," **Sean Whalen** and Katherine S. Pollard, *Genome Research.*

2018   "The Epstein-Barr virus episome maneuvers between nuclear chromatin compartments during reactivation," Stephanie Moquin, Sean Thomas, **Sean Whalen** et al., *Journal of Virology.*

2017   "Analysis of Transcriptional Variability in a Large Human iPSC Library Reveals Genetic and Non-genetic Determinants of Heterogeneity," Ivan Carcamo-Orive et al., *Cell Stem Cell.*

2017   "Genomic analyses for age at menarche identify 389 independent signals and indicate BMI-independent effects of puberty timing on cancer susceptibility," Felix R. Day et al., *Nature Genetics.*

2016   "Enhancer-promoter interactions are encoded by complex genomic signatures on looping chromatin," **Sean Whalen**, Rebecca M. Truty, and Katherine S. Pollard, *Nature Genetics.*

2016   "Unboxing Cluster Heatmaps," Sophie Engle, **Sean Whalen**, Alark Joshi, and Katherine S. Pollard, *Proceedings of the 6th Symposium on Biological Data Visualization (BioVis).*

2015   "Prediction of Human Population Responses to Toxic Compounds by a Collaborative Competition," Federica Eduati et al., *Nature Biotechnology.*

2015   "Predicting Protein Function and Other Biomedical Characteristics with Heterogeneous Ensembles," **Sean Whalen**, Om Pandey, and Gaurav Pandey, *Methods.*

2014   "Model Aggregation for Distributed Content Anomaly Detection," **Sean Whalen**, Nathaniel Boggs, and Salvatore J. Stolfo, *Proceedings of the 7th ACM Workshop on Artificial Intelligence and Security.*

2014   "Enhancing the Functional Content of Eukaryotic Protein Interaction Networks," Gaurav Pandey, Sonali Arora, Sahil Manocha, and **Sean Whalen**, *PLoS ONE.*

2013   "A Comparative Analysis of Ensemble Classifiers: Case Studies in Genomics," **Sean Whalen** and Gaurav Pandey, *Proceedings of the 13th IEEE International Conference on Data Mining.*

2013   "Multiclass Classification of Distributed Memory Parallel Computations," **Sean Whalen**, Sean Peisert, and Matt Bishop, *Pattern Recognition Letters.*

2012   "Visualizing Distributed Memory Computations with Hive Plots," Sophie Engle and **Sean Whalen** (co-first), *Proceedings of the 9th International Symposium on Visualization for Cyber Security.*

2012   "Structural Drift: The Population Dynamics of Sequential Learning," James P. Crutchfield and **Sean Whalen** (co-first), *PLoS Computational Biology.*

2012   "Network-Theoretic Classification of Parallel Computation Patterns," **Sean Whalen**, Sophie Engle, Sean Peisert, and Matt Bishop, *International Journal of High Performance Computing Applications.*

2012   "A Taxonomy of Buffer Overflow Characteristics," Matt Bishop, Sophie Engle, Damien Howard, and **Sean Whalen**, *IEEE Transactions on Dependable and Secure Computing.*

2011   "This is the Remix: Structural Improvisation using Automated Pattern Discovery," **Sean Whalen** and James P. Crutchfield, *Proceedings of the 4th International Workshop on Machine Learning and Music.*

2011	"Network-Theoretic Classification of Parallel Computation Patterns," **Sean Whalen**, Sean Peisert, and Matt Bishop, *Proceedings of the 1st International Workshop on Characterizing Applications for Heterogeneous Exascale Systems.*
2010	"Hidden Markov Models for Automated Protocol Learning," **Sean Whalen**, Matt Bishop, and James P. Crutchfield, *Proceedings of the 6th International ICST Conference on Security and Privacy in Communication Networks.*
2009	"Case Studies of an Insider Framework," Matt Bishop, Sophie Engle, Carrie Gates, Sean Peisert, and **Sean Whalen**, *Proceedings of the 42nd Annual Hawaii International Conference on System Sciences.*
2008	"We Have Met the Enemy and He is Us," Matt Bishop, Sophie Engle, Carrie Gates, Sean Peisert, and **Sean Whalen**, *Proceedings of the 2008 New Security Paradigms Workshop.*

Book Chapters

2010	"A Risk Management Approach to the Insider Threat," Matt Bishop, Sophie Engle, Deborah A. Frincke, Carrie Gates, Frank L. Greitzer, Sean Peisert, and **Sean Whalen**, in "Insider Threats in Cybersecurity—And Beyond," Springer Verlag, Berlin.

Talks & Posters

2016	"Enhancer-promoter interactions are encoded by complex genomic signatures on looping chromatin", Biology of Genomes 2016, Cold Spring Harbor, NY.
2016	"Supervised ensembles to boost the predictive power of DREAM challenges", Intelligent Systems for Molecular Biology 2016, Orlando, FL.
2014	"The Genesips Project: an NHLBI-Sponsored induced Pluripotent Stem Cell (iPSC) Resource for the Study of Cardiovascular Diseases", American Heart Association Scientific Session 2014, Chicago, IL.
2014	"Predicting Protein Function with Heterogeneous Ensembles", Automated Protein Function Prediction 2014 (held in conjunction with the 22nd International Conference on Intelligent Systems in Molecular Biology), Boston, MA.
2014	"Enhancing the Functional Content of Protein Interaction Networks", Poster, Automated Protein Function Prediction 2014 (held in conjunction with the 22nd International Conference on Intelligent Systems in Molecular Biology), Boston, MA.
2012	"Measuring Access, Knowledge, and Trust: A Discussion", Dagstuhl Workshop on Organizational Processes for Supporting Sustainable Security, Dagstuhl, Germany.
2011	"This is the Remix: Structural Improvisation using Automated Pattern Discovery", 4th International Workshop on Machine Learning and Music (held in conjunction with the 25th Annual Conference on Neural Information Processing Systems), Sierra Nevada, Spain.
2011	"Hybrid Approaches for Classifying Parallel Computation," Poster, Department of Energy Applied Mathematics Meeting, Washington DC.
2011	"Network-Theoretic Classification of Parallel Computation Patterns", 1st International Workshop on Characterizing Applications for Heterogeneous Exascale Systems (held in conjunction with the 25th International Conference on Supercomputing), Tucson, AZ.
2010	"Hidden Markov Models for Automated Protocol Learning", 6th International Conference on Security and Privacy in Communication Networks, Singapore.

2010 "Anomaly Detection in High Performance Computing", Discovery 2015: Cloud Computing Institute, UCBerkeley, Berkeley, CA.

2008 "Realtime Social Network Analysis", SFI Workshop on Immersive Visualization, Santa Fe Institute, Santa Fe, NM.

Service

Additional Teaching

2017 3 Day Intensive Python Training (fundamentals/numpy/pandas/scikit-learn), USF Data Science Institute

2016-2017 2-Part Lab, Intro to Machine Learning with R, Systems Pharmacology PSPG245, UCSF, Sourav Bandyopadhyay & Rada Savic

2015 Lab, Intro to Machine Learning with R, Statistical Methods for Bioinformatics BMI206, UCSF, Katie Pollard

2015 Lab, Intro to Machine Learning with R, Systems Pharmacology PSPG245, UCSF, Natalia Khuri

2014 Lab, Intro to Machine Learning with Python, Statistical Methods for Bioinformatics BMI206, UCSF, Katie Pollard

Committees

2016 Program Committee, 13th International Symposium on Visualization for Cyber Security

2015 Program Committee, 12th International Symposium on Visualization for Cyber Security

2015 Organizing Committee, 36th IEEE Symposium on Security and Privacy

2015 Program Committee, Cyber and Information Security Research Conference

2014 Program Committee, Cyber Security Awareness Week

2014 Program Committee, Automated Protein Function Prediction 2014

2014 Program Committee, 11th International Symposium on Visualization for Cyber Security

2014 Organizing Committee, 35th IEEE Symposium on Security and Privacy

2013 Program Committee, Cyber Security Awareness Week

2012 Program Committee, 5th Workshop on Cyber Security Experimentation and Test

Reviews

12th & 13th Annual IEEE Conference on Technologies for Homeland Security; 19th & 20th ACM SIGKDD Conference on Knowledge Discovery and Data Mining; 22nd ACM International Conference on Information and Knowledge Management; 5th European Workshop on Systems Security (ACM SIGOPS EuroSys Conference); Bioinformatics; Chromosome Research; IEEE Computer; IEEE Internet Computing; IEEE Transactions on Dependable and Secure Computing; IEEE Transactions on Systems, Man, and Cybernetics Part B; International Journal of Information Security; Journal of Applied Mathematics; Nature Genetics; Nucleic Acids Research; R Journal

References (Alphabetical)

Prof. Matt Bishop   bishop at cs.ucdavis.edu
Prof. James P. Crutchfield   crutchfield at ucdavis.edu
Prof. Gaurav Pandey   gaurav.pandey at mssm.edu
Prof. Sean Peisert   sppeisert at lbl.gov
Prof. Katherine S. Pollard   katherine.pollard at gladstone.ucsf.edu
Prof. Salvatore J. Stolfo   sal at cs.columbia.edu