

# GENETICS, BIOINFORMATICS, AND SYSTEMS BIOLOGY COLLOQUIUM

THURSDAY MAY 11

12:00PM PST

LIVE ON ZOOM!

[CLICK HERE](#)

PASSWORD: GENOMICS

PRESENTED BY:



For information on upcoming lectures, visit [genomic.weebly.com](http://genomic.weebly.com)

SPONSORED BY:



## KATHERINE S. POLLARD, PHD

DIRECTOR, GLADSTONE INSTITUTE OF DATA SCIENCE & BIOTECHNOLOGY  
PROFESSOR, UC SAN FRANCISCO  
INVESTIGATOR, CHAN ZUCKERBERG BIOHUB

### DEEP-LEARNING DISCOVERY OF VARIANTS THAT ALTER 3D GENOME FOLDING

The human genome sequence folds in three dimensions (3D) into a rich variety of locus-specific contact patterns. Despite growing appreciation for the importance of 3D genome folding in evolution and disease, we lack models for relating mutations in genome sequences to changes in genome structure and function. Towards that goal, we developed a computational model (a deep convolutional neural network called Akita) that accurately predicts genome folding from DNA sequence alone. Representations learned by Akita underscore the importance of the structural protein CTCF but also reveal a complex grammar beyond CTCF binding sites that underlies genome folding, including an unexpectedly large contribution of repetitive elements. Akita enables rapid in silico predictions for effects of sequence variants on the 3D genome, including differences in genome folding across species and in disease cohorts. We validated high-scoring causal variants from patients with autism, developmental delay, and congenital heart defects using CRISPR-edited genomes. This prediction-first strategy exemplifies my vision for a more proactive, rather than reactive, role for data science in biomedical research.

Organization Committee: J. Gleeson, J. Sebat & BISB PhD Students  
GBSBC Seminar Coordinator: W. Harabedian  
BISB Seminar Coordinator: F. Perez Estrada

[WWW.GENOMIC.WEEBLY.COM](http://WWW.GENOMIC.WEEBLY.COM)