BIOGRAPHICAL SKETCH as of 7/2021

NAME: **Gill Bejerano**

POSITION TITLE: Professor of Developmental Biology, of Computer Science,
of Pediatrics (Medical Genetics) and of Biomedical Data Science, Stanford University

EDUCATION/TRAINING

INSTITUTION AND LOCATION	DEGREE *(if applicable)*	Completion Date	FIELD OF STUDY
Hebrew University, Jerusalem, Israel	B.Sc.	06/1997	Physics, Mathematics & Computer Science *summa cum laude*
Hebrew University, Jerusalem, Israel	Ph.D.	04/2004	Computer Science (Machine Learning & Bioinformatics)
University of California, Santa Cruz CA	Postdoc	02/2007	Computational & Functional Genomics

A. PERSONAL STATEMENT

I hold 3 major quantitative undergraduate degrees, a PhD in Machine Learning and Computational Biology, and postdoc experience in an HHMI lab that combined computational and experimental genomics & genetics. My lab is shared between Stanford Schools of Medicine and Engineering, with a primary appointment in 4 departments: Developmental Biology, Computer Science, Pediatrics (Medical Genetics) and Biomedical Data Science. My lab combines computational and experimental biology with the goal of integrating computational hypothesis generation with experimental data collection & validation and translational research. Members of my lab have diverse backgrounds that range from molecular biology, neurobiology and medicine to computer science, physics and engineering. At Stanford I have trained students from the Computer Science, Developmental Biology, Genetics, BioMedical Informatics, MD/PhD, Electrical Engineering, Computational Mathematics and Physics departments, and have nurtured postdocs who are now faculty in the U.S., Canada and Europe. These experiences have prepared me to perform research, mentor students and to interact with colleagues of all backgrounds at the intersection of biomedicine and computation.

B. POSITIONS AND HONORS

PROFESSIONAL EXPERIENCE

1996-1999 Lecturer, Computer Science Department, Open University, Israel
1997-2000 Lecturer, Jerusalem Center for Professional Computer Studies, Open University, Israel
1999-2001 Lecturer, Computer Science Department, Hadassah Academic College, Israel
2000-2003 Senior Lecturer, Software Eng. Dept., Jerusalem Academic College of Engineering, Israel
2003-2007 Postdoctoral Fellow, School of Engineering, University of California, Santa Cruz
Lab of HHMI investigator David Haussler
2007-2014 Assistant Professor, Dept. of Developmental Biology, School of Medicine, Stanford University
2007-2014 Assistant Professor, Dept. of Computer Science, School of Engineering, Stanford University
2014-2019 Associate Professor, Dept. of Developmental Biology, School of Medicine, Stanford University
2014-2019 Associate Professor, Dept. of Computer Science, School of Engineering, Stanford University
2014-2019 Associate Professor, Dept. of Pediatrics (Medical Genetics), School of Medicine, Stanford Uni.
2018-2019 Associate Professor, Dept. of Biomedical Data Science, School of Medicine, Stanford University
2020- Professor, Dept. of Developmental Biology, School of Medicine, Stanford University
2020- Professor, Dept. of Computer Science, School of Engineering, Stanford University
2020- Professor, Dept. of Pediatrics (Medical Genetics), School of Medicine, Stanford Uni.
2020- Professor, Dept. of Biomedical Data Science, School of Medicine, Stanford University

HONORS AND AWARDS

1993-1996 Rector Prize & Dean's list for undergraduate achievements, Hebrew University
1996 Intel award for achievements
1997-1999 Rector Prize & Dean's list for graduate studies achievements, Hebrew University
1999 Rachel & Salim Banin scholarship

1999-2002	Levi Eshkol graduate studies fellowship
1999,2003	Best paper by a young scientist award, RECOMB conference
2007	U.S. National Academy of Science Kavli symposium invitee
2007-2010	Junior Faculty Grant, Edward Mallinckrodt, Jr. Foundation
2008	Tomorrow's Principal Investigator, Genome Technology magazine
2008-2010	Research Fellow, Alfred P. Sloan Foundation
2008-2011	Young Investigator Award, Human Frontier Science Program
2008-2011	Searle Scholar, Kinship Foundation Searle Scholars Program
2008-2009	Research Grant Award, Okawa Foundation
2008-2013	Fellowship for Science and Engineering, David and Lucile Packard Foundation
2008	UC Berkeley Center for Computational Biology Retreat keynote
2009	Microsoft Faculty Fellow
2010	U. Chicago Committee on Genetics, Genomics & Systems Biology student invited speaker
2011	Pinkham Basic Sciences Lectureships invited speaker
2012	Weizmann Institute student organized Evo Devo conference invited speaker
2013	Broad Institute Conference keynote speaker
2014-2018	Sony Faculty Scholar Award
2015-	Gene Ontology Consortium Scientific Advisory Board
2016	Keystone meeting keynote speaker
2018	Best paper award, ASPLOS conference
2018	Boston Evolutionary Genomics Retreat keynote speaker
2019	Weizmann Institute student organized Evolution conference opening keynote speaker
2019-	NIH/NHGRI Genomic Data Science Working Group Member
2019	Israel Science Foundation, Israel Precision Medicine Program evaluation committee
2019	i-Core Chromatin and RNA meeting keynote speaker
2019	U. Chicago Department of Genetics Postdoc invited speaker

C. CONTRIBUTIONS TO SCIENCE. (not including commercial, administrative and training)

1) Towards the automation of Mendelian patient diagnosis. Since joining the Stanford department of Pediatrics (1/2014) we have used our genomic expertise for patient genome diagnosis. We were the first to alert the biomedical community to the value of Mendelian patient exome reanalysis [1], a result that has since been replicated by many, leading to the introduction of a new reanalysis billing code. We developed M-CAP, the first clinically-oriented nonsynonymous variant pathogenicity predictor [2], and later S-CAP the first splicing pathogenicity predictor (*Nature Genetics*, 2019). We launched the AMELIE [3] portal to analyze PubMed literature daily in search of novel patient diagnoses. Most recently we developed InpherNet [4] to improve the prediction of novel disease genes from neighboring genes in a large graph of paralogs and orthologs.

1. Wenger AM, Guturu H, Bernstein JA & **Bejerano G**. Systematic reanalysis of clinical exome data yields additional diagnoses: implications for providers. **Genet. Med.** 19(2):209-214, 2017, e-pub 2016.
2. K.A. Jagadeesh, A.M. Wenger, M.J. Berger, H. Guturu, P.D. Stenson, D.N. Cooper, J.A. Bernstein and **G. Bejerano**. M-CAP eliminates a majority of variants of uncertain significance in clinical exomes at high sensitivity. Nature Genetics, 48(12):1581-1586, 2016. (cited 453 times)
3. J. Birgmeier, M. Haeussler, C.A. Deisseroth, E.H. Steinberg, K.A. Jagadeesh, Al.J. Ratner, H. Guturu, A.M. Wenger, M.E. Diekhans, P.D. Stenson, D.N. Cooper, C. Ré, The Manton Center, A.H. Beggs, J.A. Bernstein, and **G. Bejerano**. AMELIE speeds Mendelian diagnosis by matching patient phenotype and genotype to primary literature. **Science Transl. Med.** 12(544):eaau9113, 2020.
   Companion web portal https://amelie.stanford.edu/ has served **over 100 thousand job submissions**.
4. B. Yoo, J. Birgmeier, J.A. Bernstein and **G. Bejerano**. InpherNet accelerates monogenic disease diagnosis using patients' candidate genes' neighbors. **Genet. Med.**, e-pub, 2021.

2) Cryptographic methods in the service of genomic and genetic privacy. Our genome encodes very rich information about our physical and mental traits, and disease susceptibilities. Many critical computations can be performed using a person's genome. Cryptography, on the other hand, has developed a set of methods to perform exact computations without the different parties sharing their inputs to the process. We have adapted these techniques in the following: Work with patient genomes across different hospital silos to discover causal

variants and novel disease genes [1]; Combine genomic and phenotypic data across hospital silos to build patient cohorts and provide third party diagnoses [2]; Allow law enforcement agencies to check the genetic profile of a potentially innocent suspect against large criminal databases such that if no match is found the exonerated suspect profile can be disposed of in the field and their anonymity protected [3].

1. K.A. Jagadeesh, D.J. Wu, J.A. Birgmeier, D. Boneh and **G. Bejerano**. Deriving genomic diagnoses without revealing patient genomes. **Science**, 357(6352):692-695, 2017.
2. K.A. Jagadeesh, D.J. Wu, J. Birgmeier, D. Boneh, and **G. Bejerano**. Keeping patient phenotypes and genotypes private while seeking disease diagnoses **bioRxiv**, https://doi.org/10.1101/746230
3. J.A. Blindenbach, K.A. Jagadeesh, **G. Bejerano***, D.J. Wu*. Avoiding Genetic Racial Profiling in Criminal DNA Profile Databases. **Nature Computational Science**, 1(4):272–279, 2021. *Co-corresponding authors. Cover story.

3) The surprising evolution of gene regulation in the human genome. The study of function in the human genome used to be restricted to genes and proximal promoters. We were the first to show that the majority of the most evolutionarily “Ultraconserved” regions in the human genome are in fact non-coding loci, which serve as developmental enhancers of nearby key developmental genes [1]. We then discovered that an exonic ultra-conserved element originated in a hitherto unknown repeat family which is still active in the coelacanth genome. In tetrapods, it died long ago and left behind a surprising set of co-opted developmental enhancers and protein coding exons [2]. Work to understand ultraconservation continues in labs around the world. This insight encouraged us to develop methodology that highlights conserved non-coding elements (CNEs) in the human genome, and to offer it (via the UCSC genome browser) as a dataset that has become a standard for the field [3]. In zebrafish, a widely used vertebrate model organism, the predominant thought was that only genes can be studied there to inform us of the human condition. We developed a computational approach to show many thousands of developmental enhancers are orthologously conserved in zebrafish, and can be studied there to reveal enhancer contribution to human health [4].

1. **G. Bejerano**, M. Pheasant, I.V. Makunin, S. Stephen, W.J. Kent, J.S. Mattick and D. Haussler. Ultraconserved elements in the human genome. **Science**, 304(5675): 1321-1325, 2004. (cited 1,839 times)
2. **G. Bejerano**, C.B. Lowe, N. Ahituv, B. King, A. Siepel, S.R. Salama, E.M. Rubin, W.J. Kent and D. Haussler. A Distal Enhancer and an Ultraconserved Exon are Derived From a Novel Retroposon. **Nature**, 441:87-90, 2006. (cited 528 times)
3. A. Siepel, **G. Bejerano**, J.S. Pedersen, A. Hinrichs, M. Hou, K. Rosenbloom, H. Clawson, J. Spieth, L.W. Hillier, S. Richards, G.M. Weinstock, R.K. Wilson, R.A. Gibbs, W.J. Kent, W. Miller, and D. Haussler. Evolutionarily Conserved Elements in Vertebrate, Fly, Worm, and Yeast Genomes. **Genome Research**, 15(8):1034-1050, 2005. (cited 3,826 times)
4. R. Madelaine, J.H Notwell, G. Skariah, C. Halluin, C.C Chen, **G. Bejerano** and P. Mourrain. A screen for deeply conserved non-coding GWAS SNPs uncovers a MIR-9-2 functional mutation associated to retinal vasculature defects in human. **Nucleic Acids Res.,** 46(7):3517-3531, 2018.

4) The functional interpretation of the gene regulatory landscape of the human genome. When the gene regulatory landscape of a cell population is measured (using ChIP-seq, DNAse-seq etc.), researchers would convert the list of proximal peaks obtained into a gene list, which was analyzed using enrichment using tools developed for microarrays. We showed that in fact peak lists should not be converted into gene lists. We developed a novel approach that evaluated distal binding events, multiple binding events next to the same gene and more to obtain a novel enrichment test that accurately measures cis regulatory enrichment. We also implemented this novel approach as a web portal, called GREAT [1]. The sequencing revolution has exposed a landscape where each transcription factor binds many thousands of sites in each cellular context, but affects only a fraction of the nearby genes. We perfected methods to predict conserved binding sites that have persisted over tens of millions of years of evolution and are more likely to function and alter fitness measurably when mutated. We also showed that if we apply GREAT to these conserved sites of a transcription factor we could predict transcription factor function [2]. We later expanded this approach to accurately predict rigid transcription factor dimers [3] and even predict human medical histories [4] from their genome-wide signatures.

1. C.Y. McLean, D. Bristor, M. Hiller, S.L. Clarke, B.T. Schaar, C.B. Lowe, A.M. Wenger and **G. Bejerano**. GREAT improves functional interpretation of cis-regulatory regions. **Nature**

Biotechnology, 28(5):495-501, 2010. (cited 2,901 times)
Companion web portal http://great.stanford.edu/ has served **over 4 million job submissions**.

2. A.M. Wenger, S.L. Clarke, H. Guturu, J. Chen, B.T. Schaar, C.Y. McLean and **G. Bejerano**. PRISM offers a comprehensive genomic approach to transcription factor function prediction. **Genome Research**, 23(5):889-904., 2013.

3. H. Guturu, A.C. Doxey, A.M. Wenger and **G. Bejerano**, Structure-aided prediction of mammalian transcription factor complexes in conserved non-coding elements. **Philos. Trans. R. Soc. B** 368(1632):20130029, 2013.

4. H. Guturu, S. Chinchali, S.L. Clarke and **G. Bejerano**, Erosion of Conserved Binding Sites in Personal Genomes Points to Medical Histories. **PLoS Computational Biology**, 12(2): e1004711, 2016.

5) The contribution of functional genomic region inactivation to evolution and disease. Throughout evolution species gain novel traits and shed ancestral ones. We discovered developmental enhancers conserved between chimpanzee and other primates and mammals, yet uniquely lost in all humans [1]. We later developed a “forward genomics” paradigm to link genes to the genomic traits they may encode, by matching the patterns of trait preservation and loss (Hiller, 2012). We recently expanded our automated approach to any gene clearly lost in at least two related species [4]. We also expanded the functional elements we screen into the non-coding genome [2], and extended our scope from independent losses to convergent evolution [3].

1. C.Y. McLean, P.L. Reno, A.A. Pollen, A.I. Bassan, T.D. Capellini, C. Guenther, V.B. Indjeian, X. Lim, D.B. Menke, B.T. Schaar, A.M. Wenger, **G. Bejerano*** and D.M. Kingsley*. Human-specific loss of regulatory DNA and the evolution of human-specific traits. **Nature**, 471:216-219, 2011. *co-corresponding authors. (cited 455 times)

2. M.J. Berger, A.M. Wenger, H. Guturu and **G. Bejerano**. Independent erosion of conserved transcription factor binding sites points to shared hindlimb, vision and external testes loss in different mammals. **Nucleic Acids Research**, 12;46(18):9299-9308, 2018.

3. A. Marcovitz*, Y. Turakhia*, H.I. Chen*, M. Gloudemans, B.A. Braun, H. Wang, and **G. Bejerano**. A functional enrichment test for molecular convergent evolution finds a clear protein-coding signal in echolocating bats and whales. **Proc. Nat’l Acad. Sci. USA**, 116(42):21094-21103, 2019.

4. Y. Turakhia*, H.I. Chen*, A. Marcovitz* and **G. Bejerano**. A fully-automated method discovers loss of mouse-lethal and human-monogenic disease genes in 58 mammals. **Nucleic Acids Res.** 48(16):e91, 2020.

Most of my published works can be found at:
Biomedical https://www.ncbi.nlm.nih.gov/pubmed/?term=%22bejerano-g%22%5Bau%5D
Computer Science https://dblp2.uni-trier.de/pers/hd/b/Bejerano:Gill

Selected preprints in review/revision:
1. Y. Tanigawa, E.S. Dyer, and **G. Bejerano**. WhichTF is dominant in your open chromatin data? **bioRxiv**, https://doi.org/10.1101/730200

2. J.K. Schull*, Y. Turakhia*, W.J. Dally, and **G. Bejerano**. Champagne: Whole-genome phylogenomic character matrix method places Myomorpha basal in Rodentia. **bioRxiv**, https://doi.org/10.1101/803957

3. Johannes Birgmeier, Ethan Steinberg, Ethan E. Bodle, Cole A. Deisseroth, Karthik A. Jagadeesh, Jennefer N. Kohler, Devon Bonner, Shruti Marwaha, Julian A. Martinez-Agosto, Stan Nelson, Christina G. Palmer, Joy D. Cogan, Rizwan Hamid, Joan M. Stoler, Joel B. Krier, Jill A. Rosenfeld, Paolo Moretti, David R. Adams, Vandana Shashi, Elizabeth A. Worthey, Christine M. Eng, Euan A. Ashley, Matthew T. Wheeler, Undiagnosed Diseases Network, Peter D. Stenson, David N. Cooper, Jonathan A. Bernstein, Gill Bejerano. AMELIE 3: Fully Automated Mendelian Patient Reanalysis at Under 1 Alert per Patient per Year. medRxiv https://doi.org/10.1101/2020.12.29.20248974

4. David Wei Wu, Jonathan A. Bernstein, Gill Bejerano Discovering Monogenic Patients with a Confirmed Molecular Diagnosis in Millions of Clinical Notes with MonoMiner medRxiv https://doi.org/10.1101/2021.07.05.21259995