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: Cherry, J. Michael

eRA COMMONS USER NAME (credential, e.g., agency login): cherry.mike

POSITION TITLE: Professor (Research) of Genetics, Emeritus

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
Purdue University, West Lafayette, Indiana	B.S.	05/1979	Biological Sciences
Purdue University, West Lafayette, Indiana	B.S.	05/1979	Biochemistry
University of California, Berkeley	Ph.D.	10/1985	Molecular Biology
Harvard University, Cambridge, Massachusetts	Postdoc	1985-1988	Genetics

A. Personal Statement

- The primary goal of my research is the integration of biological knowledge and experimental results into software and database environments that provide access to these data with sophisticated tools. My laboratory also focuses on developing ontologies that unify biological nomenclature by aiding curation and computation, and the subsequent public distribution of this knowledge. I direct a group of Scientific Biocurators, Software Engineers, Data Wranglers, and Bioinformatic Analysts who serve the biological research and teaching communities through our public resources. I have been active in bioinformatics since 1990 and have been the Director of the Saccharomyces Genome Database (SGD), a major model organism database, since 1993. This project focuses on expert curation and synergizing published results from both focused single-gene studies and genomic-scale projects. We optimize the curatorial workflow using text mining of full-text peer-reviewed papers and the focused identification of laboratory methods typically associated with determining gene product function. The captured data are combined in a database used by skilled biocurators and verified by bioinformatic analysts. The assimilated knowledge includes gene function, regulation, interactions, expression, cellular and metabolic pathways, evolutionary relationships, molecular modifications, and observations from high-throughput genomic analyses. Open-source computational tools are provided to assist researchers in their exploration leading to discovery. I was the Principal Investigator for the NIH ENCODE Data Coordination Center (DCC), where our research focuses on defining and applying rigorous methods to manage the results of thousands of epigenomic assays and distributing these results in an open and effective manner. These datasets are integrated via rich metadata into a system developed and distributed publicly via the ENCODE Portal, developed by my group. I am now PI of the Data Administration and Coordination Center (DACC) for the Impacts of Genomic Variation on Function (IGVF) consortium. We have expanded on our experience creating and maintaining the ENCODE DCC to this new large consortium focusing on large scale genomic perturbations and the molecular phenotypes that are observed via techniques such as through single-cell RNA-seq. I am also one of the MPIs for the Gene Ontology Consortium (GOC), which has created and maintains the standard language to describe protein and RNA function. As part of the GOC, my lab focuses on developing curation processes, defining the consistency and reproducibility of the resulting annotations, and educating on the use and application of the ontologies and annotations. As the contact PI for the RegulomeDB project, I oversee a resource that provides integrated knowledge of the wealth of existing information concerning regulatory elements within non-exonic regions of the genome. My group has recently joined the Human Cell Atlas (HCA) project, which aims to create reference maps of human cells using single-cell omics and imaging assays. The laboratory provides the Data Operations and Curation component that coordinates the metadata, data wrangling, and outreach of the CZI Seed Network component of the HCA.

Ongoing projects that I would like to highlight include:

U24 HG001315
Cherry (PI)
03/10/2021 – 02/28/2026
Genomic Resource for the Yeast Saccharomyces

U24 HG012012
Cherry (PI)
09/01/2021 - 05/31/2026
A Data and Administrative Coordinating Center for the Impact of Genomic Variation on Function Consortium.

U24 HG010859
Sternberg (PI), Role: subcontract PI
09/18/2019 - 07/31/2024
Alliance Central: A Platform for Sustainable development of next generation genome knowledgebases

U24 HG009293
Cherry (PI)
09/01/2020 - 06/30/2025
RegulomeDB: A Resource for the Human Regulome

U24 HG012212
Thomas (PI), Role: subcontract-PI
03/03/2022 - 02/28/2027
Gene Ontology Consortium and Knowledgebase

CZIF2022-007544
Cherry (PI)
06/01/2022 - 05/31/2025
Lattice: Data Curation & Wrangling for CZI Single Cell Biology

B. Positions, Scientific Appointments, and Honors

Positions and Scientific Appointments

2024 –	Professor (Research) of Genetics, Emeritus, Stanford University (status Active)
2013 – 2023	Professor (Research) of Genetics, Stanford University
2001 – 2013	Associate Professor (Research) of Genetics, Stanford University
1998 – 2000	Director, Stanford Microarray Database, Stanford University, NIH grant to Stanford University (co-PIs: Pat Brown & David Botstein)
1993 – 2001	Chief Curator & Director, *Saccharomyces* Genome Database, NIH grant to Stanford University (PI: David Botstein)
1993 – 1996	Head, Computing. Stanford DNA Sequence & Technology Center, NIH grant to Stanford University (PI: Ron Davis)
1991 – 1993	Project Manager, AAtDB (An *Arabidopsis thaliana* Database). Massachusetts General Hospital & Department of Genetics, Harvard University (PI: Howard Goodman)
1991 – 1993	Research Associate in Genetics, Harvard University
1988 – 1993	Director of Computing, Molecular Biology, Massachusetts General Hospital
1985 – 1988	Research Fellow in Genetics, Harvard University

Other Experience and Professional Memberships

2020 –	Scientific Advocate/Advisory Board, GlyGen, The George Washington University, PI: Raja Mazumder
2016 –	Scientific Advisory Board, XenBase, University of Calgary, Canada, NIH NICHD, PI: P. Vize
2015 – 2017	President, International Society of Biocuration
2015 –	External Scientists Panel, FaceBase, NIH NIDCR, University of Southern California
2012 – 2023	Member, Stanford Center of Computational, Evolutionary, and Human Genomics (CEHG)
2011 – 2023	Associate Editor, G3: Genes | Genomes | Genetics
2010 – 2023	Associate Editor, Database - The Journal of Biological Databases and Curation
2010 – 2017	Executive Board, International Society of Biocuration, re-elected in 2012 & 2014
2009 –	Member, Stanford Center for Genomics and Personalized Medicine (SCGPM)
2008 – 2018	Member Advisory Committee, FlyBase, *Drosophila* database, Harvard, PI: Norbert Perrimon
2002 – 2018	Member Advisory Committee, WormBase, *Caenorhabditis* database, Caltech PI: Paul Sternberg

Honors

2017	Genetics Society of America, Ira Herskowitz Award, presented at the Yeast Genetics and Molecular Biology Meeting at Stanford University

C. Contributions to Science

My research focuses on identifying, validating, and integrating scientific information into encyclopedic databases essential for both investigation and scientific education. Published results of scientific experimentation form the foundation of our understanding of the natural world and provide motivation for new experiments. Combining the in-depth understanding reported in the literature with computational analyses is a crucial aspect of modern biological research. Mastering the vast volumes of published literature requires comprehensive databases that present the facts and underlying experimental data in publicly accessible ways. Curation, extraction, and sorting of factual experimental data from peer-reviewed journal articles are necessary to acquire these data from their sources. Large quantitative datasets from genomic studies expand our knowledge of genes, their products, and their interactions. By integrating these quantitative datasets with curated experimental results, we create unique, comprehensive databases. My group develops such essential databases and makes them available to scientists and educators who seek to understand experimental results and teach scientific knowledge.

1. My work has been focused on the creation of the *Saccharomyces* Genome Database. The database of fundamental published information on the budding yeast *Saccharomyces cerevisiae*. (www.yeastgenome.org).
    a. Skrzypek, M.S., Nash, R.S., Wong, E.D., MacPherson, K.A., Hellerstedt, S.T., Engel, S.R., Karra, K., Weng, S., Sheppard, T.K., Binkley, G., Simison, M., Miyasato, S.R. and **J.M. Cherry**. (2018) Saccharomyces Genome Database informs human biology. *Nucleic Acids Res*. (Database Issue) doi: 10.1093/nar/gkx1112 PMCID: PMC5753351
    b. Wong, E.D., Skrzypek, M.S., Weng, S., Binkley, G., Meldal, B.H., Perfetto, L., Orchard, S.E., Engel, S.R., **Cherry, J.M**. (2019) Integration of macromolecular complex data into the Saccharomyces Genome Database. *Database (Oxford)*. doi: 10.1093/database/baz008 PMCID: PMC6360207
    c. Ng PC, Wong ED, MacPherson KA, Aleksander S, Argasinska J, Dunn B, Nash RS, Skrzypek MS, Gondwe F, Jha S, Karra K, Weng S, Miyasato S, Simison M, Engel SR, **Cherry JM**. Transcriptome visualization and data availability at the Saccharomyces Genome Database. (2020) *Nucleic Acids Res.* doi: 10.1093/nar/gkz892 PMID: 31612944 PMCID: PMC7061941
    d. Engel, S.R., Wong, E.D., Nash, R.S., Aleksander, S., Alexander, M., Douglass, E., Karra, K., Miyasato, S.R., Simison, M., Skrzypek, M.S., Weng, S. and **Cherry, J.M**. (2022) New data and collaborations at the Saccharomyces Genome Database: updated reference genome, alleles, and

    the Alliance of Genome Resources. Genetics. 220. doi: 10.1093/genetics/iyab224, PMID: 34897464, PMCID: PMC9209811

2. Establishment of the Alliance of Genome Resources and the Alliance Central Resource (alliancegenome.org). A ground-breaking integration of biological information that defines and manages reference data standards and application of these as an online portal.
    a. The Alliance of Genome Resources: Building a Modern Data Ecosystem for Model Organism Databases. (2019) *Genetics*. 213:1189-1196. doi: 10.1534/genetics.119.302523 PMID: 31796553; PMCID: PMC6893393
    b. Alliance of Genome Resources Portal: unified model organism research platform. (2020) *Nucleic Acids Res*. 48(D1):D650-D658. doi: 10.1093/nar/gkz813 PMID: 31552413; PMCID: PMC6943066
    c. Alliance of Genome Resources Consortium. (2022) Harmonizing model organism data in the Alliance of Genome Resources. Genetics 220 doi: 10.1093//genetics/iyac022 PMID: 35380658, PMCID: PMC8982023
    d. Alliance of Genome Resources Consortium. (2024) Updates to the Alliance of Genome Resources central infrastructure. Genetics 227 doi: 10.1093//genetics/iyae049 PMID: 38552170, PMCID: PMC11075569

3. Creation and application of ontologies for enhanced annotation of metadata for ENCODE experimental results, and the use of metadata to facilitate enhanced access to these data. (www.encodeproject.org)
    a. Malladi, V.S., Erickson, D.T., Podduturi, N.R., Rowe, L.D., Chan, E.T., Davidson, J.M., Hitz, B.C., Ho, M., Lee, B.T., Miyasato, S., Roe, G.R., Simison, M., Sloan, C.A., Strattan, J.S., Tanaka, F., Kent, W.J., **Cherry, J.M**., and E.L. Hong. (2015) Ontology application and use at the ENCODE DCC. *Database (Oxford).* doi: 10.1093/database/bav010 PMCID: PMC4360730
    b. Hong, E.L., Sloan, C.A., Chan, E.T., Davidson, J.M., Malladi, V.S., Strattan, J.S., Hitz, B.C., Gabdank, I., Narayanan, A.K., Ho, M., Lee, B.T., Rowe, L.D., Dreszer, T.R., Roe, G.R., Podduturi, N.R., Tanaka, F., Hilton, J.A., and **Cherry, J.M**. (2016) Principles of metadata organization at the ENCODE Data Coordination Center. *Database (Oxford).* doi: 10.1093/database/baw001 PMCID: PMC4792520

4. Construction of rich metadata models, robust unified computational analyses, and user portal (www.encodeproject.org) for the ENCODE Consortium.
    a. Hitz, B., Rowe, L.D., Podduturi, N., Glick, D., Baymuradov, U., Malladi, V., Chan, E., Davidson, J., Gabdank, I., Narayanan, A., Onate, K., Ho, M., Lee, B., Miyasato, S., Dreszer, T., Sloan, C., Strattan, J.S., Tanaka, F., Hong, E.L., and **Cherry, J.M**. (2016) SnoVault and encodeD: A novel object-based storage system and applications to ENCODE metadata. *PLoS One* doi: 10.1371/journal.pone.0175310 PMCID: PMC5389787
    b. Davis CA, Hitz BC, Sloan CA, Chan ET, Davidson JM, Gabdank I, Hilton JA, Jain K, Baymuradov UK, Narayanan AK, Onate KC, Graham K, Miyasato SR, Dreszer TR, Strattan JS, Jolanki O, Tanaka FY, **Cherry JM.** (2018) The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res.* 46(D1):D794-D801. doi: 10.1093/nar/gkx1081 PMID: 29126249; PMCID: PMC5753278
    c. Gabdank, I., Chan, E.T., Davidson, J.M., Hilton, J.A., Davis, C.A., Baymuradov, U.K., Narayanan, A., Onate, K.C., Graham, K., Miyasato, S.R., Dreszer, T.R., Strattan, J.S., Jolanki, O., Tanaka, F.Y., Hitz, B.C., Sloan, C.A., **Cherry, J.M**. (2018) Prevention of data duplication for high throughput sequencing repositories. *Database (Oxford).* doi: 10.1093/database/bay008 PMCID: PMC5829560
    d. Jou J, Gabdank I, Luo Y, Lin K, Sud P, Myers Z, Hilton JA, Kagda MS, Lam B, O'Neill E, Adenekan P, Graham K, Baymuradov UK, R Miyasato S, Strattan JS, Jolanki O, Lee JW, Litton C, Y Tanaka F, Hitz BC, **Cherry JM**. (2019) The ENCODE Portal as an Epigenomics Resource. *Curr Protoc Bioinformatics*. 68:e89. doi: 10.1002/cpbi.89 PMID: 31751002; NIHMSID:NIHMS1058769

5. Application of publicly available GWAS and epigenomic results to identify functional elements in human and mouse genomes, RegulomeDB (regulomedb.org).
    a. Boyle, A., Hong, E.L., Hariharan, M., Cheng, Y., Schaub, M.A., Kasowski, M., Karczewski, K.J., Park, J., Hitz, B.C., Weng, S., **Cherry, J.M**. and M. Snyder (2012) Annotation of functional variation in personal genomes using RegulomeDB. *Genome Research*. 22:1790-1797 PMCID: PMC3431494

b. Luo Y, Hitz BC, Gabdank I, Hilton JA, Kagda MS, Lam B, Myers Z, Sud P, Jou J, Lin K, Baymuradov UK, Graham K, Litton C, Miyasato SR, Strattan JS, Jolanki O, Lee JW, Tanaka FY, Adenekan P, O'Neill E, **Cherry JM**. (2019) *Nucleic Acids Res.* doi: 10.1093/nar/gkz1062 PMID: 31713622 PMCID: PMC7061942
c. Dong S, Zhao, N., Spragins, E., Kagda, M.S., Li, M., Assis, P., Jolanki, O., Luo, Y., **Cherry, J.M.,** Boyle, A.P. and Hitz, B.C. (2023) Annotating and prioritizing human non-coding variants with RegulomeDB v.2. Nat Genet. 55:724-726. doi: 10.1038/s41588-023-01365-3, PMID: 37173523,
d. Dong S, Zhao N, Spragins E, Kagda MS, Li M, Assis P, Jolanki O, Luo Y, Cherry JM, Boyle AP, Hitz BC. (2023) Annotating and prioritizing human non-coding variants with RegulomeDB v.2. Nat Genet. 55:724-726. doi: 10.1038/s41588-023-01365-3. PubMed PMID: 37173523; PubMed Central PMCID: PMC10989417.

Complete List of Published Work in MyBibliography:
https://www.ncbi.nlm.nih.gov/myncbi/joe%20michael.cherry.1/bibliography/public/