Agenda

<table>
<tr><th></th><th>Sunday
March 26, 2017</th><th>Monday
March 27, 2017</th><th>Tuesday
March 28, 2017</th><th>Wednesday
March 29, 2017</th></tr>
<tr><td>7:30 AM</td><td>Registration open</td><td>Registration open</td><td>Registration open</td><td></td></tr>
<tr><td>8:30 AM</td><td>Keynote speaker: Ami Bhatt</td><td>Keynote speaker: Michael Huerta</td><td>Keynote speaker: Daphne Koller</td><td rowspan="2">Workshops 6, 7
(Locations noted below)</td></tr>
<tr><td>9:30 AM</td><td>Session 1: Data Integration, Data Visualization, and Community Annotation</td><td>Session 3: DATABASE virtual issue</td><td>Session 6: Data Standards and Ontologies</td></tr>
<tr><td>10:30 AM</td><td>Coffee Break</td><td>Coffee break</td><td>Coffee break</td><td>Coffee break</td></tr>
<tr><td>11:00 AM</td><td>Session 1 (cont): Data Integration, Data Visualization, and Community Annotation</td><td>Session 3 (cont): DATABASE virtual issue</td><td>Session 6: Data Standards and Ontologies</td><td>Keynote speaker: Steve Lincoln</td></tr>
<tr><td>12 noon</td><td>Lunch</td><td>Lunch</td><td rowspan="2">Lunch</td><td rowspan="2">Lunch</td></tr>
<tr><td>12:30 PM</td><td rowspan="2">Poster session I, Berg Hall, Rm A</td><td rowspan="2">Poster session II, Berg Hall, Rm A</td></tr>
<tr><td>1:00 PM</td><td rowspan="4">Workshops 4, 5
(Locations noted below)</td><td rowspan="2">ISB General Meeting</td></tr>
<tr><td>1:30 PM</td><td rowspan="4">Workshops 1, 2, 3
(Locations noted below)</td><td rowspan="4">Session 4: Functional Annotation</td></tr>
<tr><td>2:00 PM</td><td>Exceptional Contributions to Biocuration Award: Chris Mungall</td></tr>
<tr><td></td><td>Biocuration Career Award: Marc Feuermann</td></tr>
<tr><td>3:00 PM</td><td>Coffee break</td><td>Coffee break</td><td>Coffee break</td></tr>
<tr><td>3:30 PM</td><td>Coffee break</td><td rowspan="2">Session 5: Text Mining</td><td>Alliance of Genome Resources</td><td rowspan="3">Session 8: Precision Medicine</td></tr>
<tr><td>4:00 PM</td><td rowspan="3">Session 2: Large Scale and Predictive Annotation/Big data</td><td>Keynote speaker: Euan Ashley</td></tr>
<tr><td>5:00 PM</td><td>Campus walking tour</td><td rowspan="3">Session 7: Curation Standards and Best Practice, Challenges in Biocuration, Biocuration Tutorial</td></tr>
<tr><td>5:30 PM</td><td rowspan="2">Cocktail reception at Stanford Faculty Club</td><td rowspan="2"></td></tr>
<tr><td>6:00 PM</td><td>Biocuration Career Award: John Westbrook</td></tr>
</table>

** All sessions will be held in Berg Hall Rooms B/C (LK 240/250) unless otherwise noted.*

Keynote Speakers

Michael Huerta, PhD
Collaborative biomedical research

Associate Director and Coordinator of Data and Open Source Initiatives

National Library of Medicine-NIH, Bethesda, Maryland

Daphne Koller, PhD
Online education as co-founder of Coursera

Chief Computing Officer

Calico Labs, South San Francisco, CA

Euan Ashley, MB ChB, MRCP, DPhil
Application of genomics and wearables to medicine

Associate Professor of Medicine (Cardiovascular), of Genetics, and of Biomedical Data Science

Stanford University, Stanford, CA

Steven Lincoln, PhD
Precision medicine

Scientific Affairs

Invitae, Palo Alto, CA

Ami Bhatt, MD, PhD
Clinical microbiome

Assistant Professor of Medicine (Hematology) and of Genetics

Stanford University, Stanford, CA

Scientific Sessions

Session 1: Data Integration, Data Visualization, and Community-based Biocuration
Sunday, March 26, 9:30 AM - 12 noon, Berg Hall Rooms B/C

Chair: Edith Wong

17. **FlyBase Gene Snapshots: e-mailing computationally predicted experts to produce short gene summaries.** *Giulia Antonazzo*, Jose-Maria Urbano and Nick H. Brown

87. **SmartAPI editor: a tool for semantic annotation of Web APIs.** *Shima Dastgheib,* Amrapali Zaveri, Trish Whetzel, Chunlei Wu and Michel Dumontier

41. **The straight mouse: defining anatomical axes in 3D embryo models.** *Chris Armit*, Bill Hill, Shanmugasundaram Venkataraman, Kenneth McLeod, Albert Burger and Richard A Baldock

43. **NaviCom: A web application to create interactive molecular network portraits using multi-level omics data.** *Inna Kuperstein,* Maturin Dorel, Eric Viara, Emmanuel Barillot and Andrei Zinovyev

36. **The Complex Portal: Broadening our horizon**. *Birgit Meldal*, Anjali Shrivastava, Colin Combe, Josh Heimbach, Maximillian Koch, Noemi Del Toro Ayllon, Henning Hermjakob and Sandra Orchard

84. **Leveraging 1,000,000 LINCS gene expression profiles to enhance curation of pharmacological mechanisms of action.** *Jodi Hirschman*, Jenny Liu, Rajiv Narayan, Mariya Khan, Ted Natoli, Bang Wong, Josh Bittker, Todd Golub, Steven Corsello and Aravind Subramanian

48. **BioMuta and BioXpress: integrated, ontology-unified databases facilitate analysis of mutation and expression landscapes across cancer with an emphasis on aberrant glycosylation in cancer.** *Hayley Dingerdissen*, Yu Hu and Raja Mazumder

85. **Repurpos.us: A fully open and expandable drug repurposing portal**. *Sebastian Burgstaller-Muehlbacher*, Núria Queralt-Rosinach, Timothy Putman, Gregory S. Stupp, Elvira Mitraka, Andra Waagmeester, Lynn Schriml, Benjamin M. Good and Andrew I. Su

Session 2: Large Scale and Predictive Annotation/Big Data
Sunday, March 26, 3:30-5:30 PM, Berg Hall Rooms B/C

Chair: Zhang Zhang

18. **Pathway and biosample mapping support hypothesis generation through visualization of nuclear receptor signaling networks in Transcriptomine.** *Lauren Becnel*, Scott Ochsner, Apollo McOwiti, Wasula Kankanamge, Alexey Naumov and Neil Mckenna

22. **The Ontology-aided biocuration in Open Targets - how biocuration pays off.** *Sirarat Sarntivijai,* Simon Jupp, Patricia Bento, Senay Kafkas, Gautier Koscielny, Barbara Palka, Gary Saunders, Ian Dunham and Helen Parkinson

39. **PedAM: A standards-based database for integrating and exchanging pediatrics-specified information from mult-level biomedical resources.** *Zhongxin An,* Jinmeng Jia, Yue Ming, Yunxiang Liang, Dongming Guo and Tieliu Shi

69. **Genome Properties at InterPro.** *Lorna Richardson*, Neil Rawlings, Gustavo Salazar-Orejuela, Alex Mitchell and Robert D. Finn

77. **Assessing Text Embedding Models for Assigning UniProt Classes to Scientific Literature.** *Douglas Teodoro*, Luc Mottin, Julien Gobeill, Cecilia Arighi and Patrick Ruch

88. **Big Data infrastructure for Chinese Human Proteome Project (CNHPP-BDI).** *Yin Huang*, Chi Jing, Yanjun Sun, Huali Xu, Yang Qiu, Jianan Zhao, Ruifeng Li, Kun Ma, Bin Li, Zhaolian Han, Jingwen Feng, Tieliu Shi, Henning Hermjakob, Jun Qin and Weimin Zhu

89**. MethBank: a DNA and RNA Methylation Databank**. *Rujiao Li*, Fang Liang, Dong Zou, Mengwei Li, Shixiang Sun and Zhang Zhang

Session 3: DATABASE Virtual Issue Session
Monday, March 27, 9:30 AM-12 noon, Berg Hall Rooms B/C

Chair: J. Michael Cherry

15. **Literature Consistency of Bioinformatics Sequence Databases is Effective for Assessing Record Quality.** *Mohamed Reda Bouadjenek*, Karin Verspoor and Justin Zobel

20. **Effective Biomedical Document Classification for Identifying Publications Relevant to the Mouse Gene Expression Database (GXD).** *Xiangying Jiang*, Martin Ringwald, Judith Blake and Hagit Shatkay

67. **Strategies towards digital and semi-automated curation in RegulonDB.** *Fabio Rinaldi*, Socorro Gama, Hilda Solano Lira, Alejandra Lopez-Fuentes, Luis José Muñiz Rascado, Cecilia Ishida-Gutiérrez, Carlos-Francisco Méndez-Cruz and Julio Collado-Vides

1. **Better living through ontologies**. *Randi Vita*, James Overton, Alessandro Sette and Bjoern Peters

73. **WikiGenomes: an open Web application for community consumption and curation of gene annotation data in Wikidata***. Timothy Putman*, Sebastien Lelong, Sebastian Burgstaller-Muehlbacher, Andra Waagmeester, Colin Diesh, Nathan Dunn, Monica Munoz-Torres, Gregory Stupp, Andrew I. Su and Benjamin Good

51. **Surveying the Maize Community for their Diversity and Pedigree Visualization Needs to Prioritize Tool Development and Curation.** *Taner Sen*, Bremen Braun, David Schott, John Portwood, Mary Schaeffer, Lisa Harper, Jack Gardiner, Ethalinda Cannon and Carson Andorf

21. **Triage by Ranking to Support the Curation of Protein Interactions.** *Luc Mottin*, Emilie Pasche, Julien Gobeill, Valentine Rech de Laval, Anne Gleizes, Pierre-André Michel, Amos Bairoch, Pascale Gaudet and Patrick Ruch

19. **Automated PDF highlights to support faster curation of literature on Parkinson's and Alzheimer's disease.** *Honghan Wu,* Anika Oellrich, Christine Girges, Bernard De Bono, Tim Jp Hubbard and Richard J. B. Dobson

62. **Curated Protein Information in the Saccharomyces Genome Database.** *Sage T. Hellerstedt*, Robert S. Nash, Shuai Weng, Kelley M. Paskov, Edith D. Wong, Kalpana Karra, Stacia R. Engel and J. Michael Cherry

74. **Outreach and online training services at the Saccharomyces Genome Database.** *Kevin A. MacPherson*, Barry Starr, Edith D. Wong, Kyla S. Dalusag, Sage T. Hellerstedt, Olivia W. Lang, Robert S. Nash, Marek S. Skrzypek, Stacia R. Engel and J. Michael Cherry

Session 4: Functional Annotation
Monday, March 27, 1:30-3:00 PM, Berg Hall Rooms B/C

Chair: Sylvain Poux

10. **EC Numbers: past, present and future.** *Ron Caspi*

23. **From laboratory to database: the *C.elegans* kinome in UniProtKB.** *Michele Magrane*, Rossana Zaru, Claire O'Donovan and Uniprot Consortium

58. **The Critical Assessment of Protein Function Annotation: The Road Ahead.** *Naihui Zhou*, Yuxiang Jiang, Timothy Bergquist, Maria J Martin, Claire O'Donovan, Sean D. Mooney, Casey S. Greene, Predrag Radivojac and Iddo Friedberg

59. **RefSeq: Curation and Annotation of Recoding Events in Vertebrates.** *Bhanu Rajput*, Terence Murphy and Kim Pruitt

61. **Automated generation of human-readable gene summaries using structured data.** *Ranjana Kishore*, James Done, Yuling Li, Juancarlos Chan, Hans Michael Muller and Paul Sternberg

98**. Using co-annotation and biological knowledge as a quality control procedure for ontology structure and gene annotation in the Gene Ontology**. *Seth Carbon*, Valerie Wood, Midori Harris, Antonia Lock, David Hill, Stacia Engel, Kimberly Vanauken and Christopher Mungall

---

Session 5: Text Mining
Monday, March 27, 3:30-5:00 PM, Berg Hall Rooms B/C

Co-chairs: Johanna McEntyre and Senay Kafkas

2. **On expert curation and sustainability: UniProtKB/Swiss-Prot as a case study**. *Sylvain Poux*, Cecilia Arighi, Michele Magrane, Zhiyong Lu and Uniprot Consortium

29. **Evaluating Automated Reading for Building Big Mechanistic Models.** *Tonia Korves,* Matthew Peterson, Christopher Garay, Robyn Kozierok and Lynette Hirschman

40. **Towards linking molecular interaction data to literature on Europe PMC.** *Aravind Venkatesan*, Senay Kafkas, Pablo Porras, Sandra Orchard and Johanna McEntyre

68. **A text mining-based approach to graph database curation in support of metabolic pathway model reconstruction.** *Riza Batista-Navarro* and Sophia Ananiadou

100. **CIViCmine: Assisting curation of the CIViC resource using relation extraction**. *Jake Lever*, Obi Griffith, Malachi Griffith and Steven Jones

28. **Integrating genomic variant information from literature with dbSNP for precision medicine.** *Zhiyong Lu*, Lon Phan and Chih-Hsuan Wei

---

Session 6: Data Standards and Ontologies
Tuesday, March 28, 9:30 AM-12 noon, Berg Hall Rooms B/C

Chair: Lynn Schriml

47. **Implementation studies for the Global Alliance for Genomics and Health data schemas.** *Michael Baudis*

44. **Biocompute objects and their potential role in evaluation and validation of HTS (NGS) computations.** *Raja Mazumder*

38. **Standardized Metadata for Mass Spectrometry-Based Proteomics.** *Yue Ming*, Jinmeng Jia, Zhongxin An, Bowen Zhong, Weimin Zhu and Tieliu Shi

66. **Genetic Interactions Structured Terminology (GIST): A new standard for describing and annotating cross-species genetic interactions data**. *Christian Grove*, Rose Oughtred, Raymond Lee, Kara Dolinski, Mike Tyers, Paul Sternberg and Anastasia Baryshnikova

86. **Development & applications of an ontology for scientific evidence, the Evidence and Conclusion Ontology (ECO).** *Marcus Chibucos*

56. **Defining genetic mechanistic subtypes in the Disease Ontology to support disease model curation and large scale data integration.** *Elvira Mitraka*, James A. Overton, Susan Bello, Stan Laulederkind, Randi Vita, Janan Eppig, Mary Shimoyama, Bjoern Peters and Lynn Schriml

49. **Challenges of ontology development for quantitative phenotype curation.** *Jennifer R. Smith*, Stan Laulederkind, Shur-Jen Wang, G. Thomas Hayman, Matthew J. Hoffman, Yiqing Zhao, Marek A. Tutaj, Jeffrey L. De Pons, Melinda R. Dwinell and Mary E. Shimoyama

---

Session 7: Curation Standards and Best Practice, Challenges in Biocuration, Biocuration Tutorial
Tuesday, March 28, 4:00-5:30 PM, Berg Hall Rooms B/C

Chair: Stacia Engel

5**. Current Issues in Biocuration.** Peter Karp

3. **Metadata Curation: The Good, the Bad and the Ugly**. *Christine Fleeman*, Kapila Patel and Anthony Chow

26. **Improving Disease Model Data Accessibility at Mouse Genome Informatics: Making the Move from OMIM to the Disease Ontology.** *Susan Bello*, Janan Eppig, Cynthia Smith and The Mgi Software Group

64. **The Variant Interpretation for Cancer Consortium: Seeking global consensus for clinical interpretation of cancer variants**. *Obi Griffith*, Malachi Griffith, David Tamborero, Alex Wagner, Kilannin Krysiak, Catherine Del Vecchio Fitz, Debyani Chakravarty, Ethan Cerami, Olivier Elemento, Nikolaus Schultz, Adam Margolin and Nuria Lopez-Bigas

76. **Creation and Implementation of Variant Curation Workflow for the ClinGen Inborn Errors in Metabolism Working Group: Phenylalanine Hydroxylase Deficiency**. *Diane B. Zastrow*, Heather Baudet, Cindy Si, Meredith A. Weaver, Angela Lager, Kristy Lee, Wei Shen, Amanda Thomas, Jonathan S. Berg, Steven F. Dobrowolski, Karen Eilbeck, Gregory Enns, Annette Feigenbaum, Uta Lichter-Konecki, Elaine Lyon, Marzia Pasquali, William J. Craigen, Rong Mao and Robert D. Steiner

79. **How open is open? An evaluation rubric for public knowledgebases.** *Melissa Haende*l, Julie McMurry and Andrew Su

---

Session 8: Curation for Precision Medicine
Wednesday, March 29, 3:30-5:30 PM, Berg Hall Rooms B/C

Chair: Jean Davidson

82. **The Monarch Initiative: Semantic data integration across species and sources for disease discovery.** *Lilly Winfree*, Julie McMurry, David Osumi-Sutherland, Damian Smedley, Chris Mungall, Melissa Haendel, Peter Robinson and Tudor Groza

37. **eRAM: encyclopedia of Rare Disease Annotation for Precision Medicine.** *Jinmeng Jia*, Zhongxin An, Yue Ming, Yunxiang Liang, Dongming Guo and Tieliu Shi

63. **Facilitating complex disease research by providing organized, accessible genetic information and analysis tools: the Type 2 Diabetes Knowledge Portal as a paradigm.** *Maria Costanzo* and Accelerating Medicines Partnership In Type 2 Diabetes

83. **CIViC: Crowdsourcing the Clinical Interpretation of Variants in Cancer.** *Kilannin Krysiak*, Nicholas Spies, Josh McMichael, Adam Coffman, Arpad Danos, Benjamin Ainscough, Cody Ramirez, Damian Rieke, Lynzey Kujan, Erica Barnell, Alex Wagner, Zachary Skidmore, Amber Wollam, Connor Liu, Martin Jones, Rachel Bilski, Robert Lesurf, Yan-Yang Feng, Nakul Shah, Melika Bonakdar, Lee Trani, Matthew Matlock, Avinash Ramu, Katie Campbell, Gregory Spies, Aaron Graubert,

Karthik Gangavarapu, James Eldred, David Larson, Jason Walker, Benjamin Good, Chunlei Wu, Andrew Su, Rodrigo Dienstmann, Adam Margolin, David Tamborero, Nuria Lopez-Bigas, Steven Jones, Ron Bose, David Spencer, Lukas Wartman, Richard Wilson, Elaine Mardis, Malachi Griffith and Obi Griffith

90. **The BIG Data Center's database resources: towards precision medicine.** *Jingfa Xiao*, Zhang Zhang, Wenming Zhao and On Behalf Of Big Data Center Members

97. **The Impact of Community Curation on Rare Disease Diagnosis.** *Ellen M. McDonagh*, Sarah Leigh, Rebecca E. Foulger, Louise Daugherty, Olivia Niblock, Maria Athanasopoulou, Alice Gardham, Arianna Tucci, Emma Baple, Chris Boustred, Andrew Devereau, Tom Fowler, Tim Hubbard, Antonio Rueda, Katherine Smith, Ellen R.A. Thomas, Clare Turnbull, Mark J. Caulfield, Richard Scott, Damian Smedley and Augusto Rendon

53. **My Cancer Genome - Precision Cancer Medicine Knowledge Resource.** *Christine Micheel*, Kathleen Mittendorf, Ingrid Anderson, Neha Jain, Michele Lenoue-Newton, Christine Lovly and Mia Levy

Posters

Session I: Sunday March 26, 2017, 12:00-1:30 PM

Berg Hall, Room A

Data Integration, Data Visualization, and Community-based Biocuration

Berg Hall, Rm A; Sunday, March 26, 12-1:30 PM

6. **Update Notifications for Biological Databases.** Suzanne Paley and Peter Karp

7. **Data integration and enrichment using Semantic Web technologies in GlyTouCan.** Kiyoko Aoki-Kinoshita, Nobuyuki Aoki, Akihiro Fujita, Noriaki Fujita, Masaaki Matsubara, Shujiro Okuda, Toshihide Shikanai, Daisuke Shinmachi, Elena Solovieva, Yoshinori Suzuki, Shinichiro Tsuchiya, Issaku Yamada and Hisashi Narimatsu

14. **PHI-base: a new interface and further additions for the multi-species pathogen–host interactions database.** Alayne Cuzick, Martin Urban, Kim Rutherford, Helder Pedro and Kim E. Hammond-Kosack

25. **It's All About the User: Employing User Driven Development Principles to Inform Design of Biological Database Interfaces and Resources**. Leonore Reiser, Tanya Berardini, Donghui Li, Qian Li, Robert Muller, Emily Strait, Andrey Vetushko and Eva Huala

33**. Micropublications: a New Way to Incentivize Community Curation and Reclaim Data Typically Inaccessible to the Science Community.** Daniela Raciti, Karen Yook, Tim Schedl, Todd Harris and Paul Sternberg

57. **Community Curation of Phenotype data in WormBase.** Christian Grove, Mary Ann Tuli, Juancarlos Chan, Karen Yook and Paul Sternberg

65. **Gramene's Plant Reactome portal: A resource for comparative plant pathway analysis.** Sushma Naithani, Justin Preece, Parul Gupta, Justin Elser, Peter D'Eustachio, Antonio Fabregat, Joel Weiser, Lincoln Stein, Doreen Ware and Pankaj Jaiswal

71. **Integrating the Clinical Interpretation of Cancer Variants with other public data in Wikidata.** Elvira Mitraka, Andra Waagmeester, Núria Queralt-Rosinach, Sebastian Burgstaller-Muehlbacher, Lynn Schriml, Josh F. McMichael, Benjamin Ainscough, Malachi Griffith, Obi L. Griffith, Andrew I. Su and Benjamin M. Good

91. **GSA: Genome Sequence Archive.** Yanqing Wang, Fuhai Song, Junwei Zhu, Sisi Zhang, Yadong Yang, Xiangdong Fang, Hongxing Lei, Zhang Zhang and Wenming Zhao

95. **Genome Warehouse.** Meili Chen, Jian Sang, Fan Wang, Wenming Zhao, Zhang Zhang and Jingfa Xiao

110. **The EMBL - European Bioinformatics Institute CRISPR Archive.** Sybilla Corbett, Thomas Juettemann, Myrto Kostadima, Fiona Cunningham, Daniel Zerbino and Paul Flicek

118. **Defining standards for the annotation and integration of disease relevant data across the model organism databases of the Alliance of Genome Resources (AGR).** Steven Marygold, Susan Bello, Yvonne Bradford, Madeline Crosby, Stacia Engel, Ranjana Kishore, Stan Laulederkind, Mary Shimoyama and Cynthia Smith

122. **Curation, processing, and data integration of information obtained via high-throughput technologies.** David Alberto Velázquez-Ramírez, Socorro Gama-Castro, Alberto Santos-Zavaleta, Mishael Sánchez-Pérez, Claire Rioualen, Cesar Bonavides-Martínez, Jacques Van Helden and Julio Collado-Vides

129. **PBD2.0: a literature-curated database for protein biomarker candidates in urine.** Chen Shao, Jingwen Guo, Lulu Zhang, Sheng Yang, Heng Wang, Jing Wei, Yongtao Liu, Na Ni, Weiwei Qin and Youhe Gao

130. **Bringing Chemical Data into FlyBase.** Silvie Fexova

137. **The i5k Workspace@NAL - a resource for arthropod genome access, visualization and community curation.** Monica F Poelchau, Mei-Ju May Chen, Yu-Yu Lin and Christopher P Childers

139. **Going Paperless: Updating Publication Acquisition and Tracking at ZFIN.** Ceri Van Slyke, Holly Paddock, Sierra Moxon, Patrick Kalita and Douglas Howe

141. **Development of an online tumor database for zoological and exotic species.** Ashley Zehnder, Tara Harrison, Cassondra Bauer, Ryan Colburn, Catherine Pfent, Joanne Paul-Murphy, Michelle Hawkins and Carlos Bustamante

145. **A Computational Framework Using Ontologies to Integrate Large-scale Trees and Traits: Exploring Diversity Across the Teleost Tree of Life.** Laura Jackson, Pasan Fernando, Josh Hanscom, James Balhoff and Paula Mabee

150. **Encouraging annotation of published works.** Christopher Hunter, Xiao Sizhe, Laurie Goodman, Peter Li and Scott Edmunds

---

Large Scale and Predictive Annotation/Big Data

Berg Hall, Rm A; Sunday, March 26, 12-1:30 PM

9. **Chemical-phenotype curation at the Comparative Toxicogenomics Database.** Allan Davis, Robin Johnson, Daniela Sciaky, Cynthia Grondin, Jolene Wiegers, Thomas Wiegers and Carolyn Mattingly

31. **The Uniprot Consortium. UniRule curation pipeline for automatic annotation of UniProtKB protein function and sequence features at the Protein Information Resource.** Qinghua Wang, Cecilia Arighi, Chuming Chen, John Garavelli, Hongzhan Huang, Kati Laiho, Darren Natale, C. R. Vinayaka, Lai-Su Yeh, Cathy Wu and The Uniprot Consortium

52. **CEDAR's Predictive Data Entry: Easier and Faster Creation of High-quality Metadata.** Marcos Martínez-Romero, Martin J. O'Connor, Ravi D. Shankar, Maryam Panahiazar, Debra Willrett, Attila L. Egyedi, Olivier Gevaert, John Graybeal and Mark A. Musen

54. **Extracting knowledge from transcriptomics big data in Bgee: integration of any dataset, including reannotation and reanalysis of GTEx, for gene list enrichment analysis, ranked gene expression patterns, and direct integration in R.** Frédéric Bastian, Anne Niknejad, Amina Echchiki, Julien Roux, Bgee Team and Marc Robinson-Rechavi.

72. **Predicting Biomedical Metadata using Rule Mining Algorithms.** Maryam Panahiazar, Michel Dumontier and Olivier Gevaert

92. **A molecular module breeding platform for rice based on a comprehensive genomic variation database.** Shuhui Song, Dongmei Tian, Cuiping Li, Dong Zou and Zhang Zhang

94. **Gene Expression Nebulas (GEN): a data portal of gene expression profiles based entirely on RNA-Seq data.** Lili Hao, Xin Sheng, Lin Xia and Zhang Zhang

103. **A machine learning method to quantify the completeness of curated data sets.** Douglas Howe

117. **Host-Pathogen Interactome: Biocuration and Computational Prediction.** Mais Ammari, Cathy Gresham, Prashanti Manda, Fiona McCarthy and Bindu Nanduri

128. **An offline-first iCLiKVAL browser extension for scientific media annotation.** Naveen Kumar and Todd Taylor

136. **Building a comprehensive catalog of Drosophila datasets at FlyBase.** Gilberto Dos-Santos, Kathleen Falls, Chris Tabone, David Emmert, Gillian Millburn, Marta Costa, Madeline Crosby and Flybase Consortium

149. **CAFA: A Community-Wide Challenge in Computational Protein Function Prediction**. Naihui Zhou, Timothy Bergquist, Yuxiang Jiang, Maria Martin, Claire O'Donovan, Sean Mooney, Casey Greene, Pedrag Radivojac and Iddo Friedberg

Functional Annotation

Berg Hall, Rm A; Sunday, March 26, 12-1:30 PM

16. **Pathway/Genome Database Editing Tools Provided By The Pathway Tools Software.** Ingrid Keseler, Suzanne Paley and Peter Karp

24. **Data curation by semantic digitization of experimental data: strengths and possibilities**. Pratibha Gour, Saurabh Raghuvanshi and Shaji Joseph

45. **Residue data and intrinsic disorder: extending InterPro functionality to improve protein sequence annotation**. Alex Mitchell, Hsin-Yu Chang, Neil Rawlings, Lorna Richardson, Amaia Sangrador and Robert D. Finn

102. **Pfam families and clans: maximizing biocuration effort**. Sara El-Gebali, Jaina Mistry, Lorna Richardson, Alex Mitchell, Alex Bateman and Rob Finn

104. **Functional annotation of proteoforms in the Mouse Genome Database using the Protein Ontology.** Harold Drabkin, Karen Christie, Cecilia Arighi, Cathy Wu and Judith Blake

108. **Integration of NCBI's Conserved Domain Database Content with InterPro.** Narmada Thanki, Shennan Lu, Farideh Chitsaz, Myra Derbyshire, Noreen Gonzales, Marc Gwadz, Fu Lu, Gabriele Marchler, James Song, Roxanne Yamashita, Chanjuan Zheng, Stephen Bryant and Aron Marchler-Bauer

112. **SPARCLE: Functional characterization of proteins by domain architecture.** Roxanne Yamashita, Aron Marchler-Bauer, Lianyi Han, Jane He, Christopher Lanczycki, Shennan Lu, Bo Yu, Farideh Chitsaz, Myra Derbyshire, Renata Geer, Noreen Gonzales, Marc Gwadz, Dave Hurwitz, Fu Lu, Gabriele Marchler, James Song, Narmada Thanki, Dachuan Zhang, Christina Zheng, Lewis Geer and Stephen Bryant

113. **Automated Generation and Optimization of Hierarchical Protein Domain Classifications for the Conserved Domain Database**. Marc Gwadz, Andrew Neuwald, Christopher Lanczycki, David Hurwitz, Farideh Chitsaz, Myra Derbyshire, Noreen Gonzales, Fu Lu, Gabriele Marchler, James Song, Narmada Thanki, Roxanne Yamashita, Chanjuan Zheng, Stephen Bryant and Aron Marchler-Bauer

120. **Comprehensive Gene Ontology annotation of ciliary genes in the laboratory mouse**. Karen R. Christie, Paola Roncaglia, Teunis J. P. van Dam, Toby J. Gibson, Jane Lomax and Judith A. Blake

134. **Xenbase: the Xenopus bioinformatics database.** Joshua Fortriede, Malcolm Fisher, Christina James-Zorn, Kevin Burns, Virgilio Ponferrada, Praneet Chaturvedi, Erik Segerdell, Kamran Karimi, Vaneet Lotay, Vicente Pader, Troy Pells, Dong Zhuo Wang, Ying Wang, Stanley Chu, Peter Vize and Aaron Zorn

138. **The ENCODE Annotation Pipeline: Standard analyses for ChIP-seq, RNA-seq, DNase-seq, and whole-genome bisulfite experiments.** J Seth Strattan, Timothy R Dreszer, Ben C Hitz, Esther T Chan, Jean M Davidson, Idan Gabdank, Jason A Hilton, Cricket A Sloan, Zhiping Weng, Anshul Kundaje, Encode Data Coordinating Center and J Michael Cherry

148. **RefSeq: Curation and Annotation of Recoding Events in Vertebrates.** Bhanu Rajput, Terence Murphy and Kim Pruitt

Session II: Monday March 27, 2017, 12:00-1:30 PM

Berg Hall, Room A

Text Mining

Berg Hall, Rm A; Monday, March 27, 12-1:30 PM

30. **Reference Set Curation for Complex Molecular Mechanisms.** Matthew Peterson, Tonia Korves, Christopher Garay, Robyn Kozierok and Lynette Hirschman

55. **Looking Under the Hood of Machine Learning for Biocuration.** Parthiban Srinivasan

70. **Author reagent table: a proposal.** Madeline Crosby, Norbert Perrimon and Flybase Consortium

106. **Recent Improvements of the BEL Information Extraction workFlow (BELIEF) for the Biomedical Text Mining and Curation.** Justyna Szostak, Marja Talikka, Juliane Fluck, Sumit Madan, William Hayes, Manuel C. Peitsch and Julia Hoeng

111. **The BioGRID Interaction Database: Curation strategies and new developments.** Lorrie Boucher, Rose Oughtred, Jennifer Rust, Christie Chang, Bobby-Joe Breitkreutz, Nadine Kolas, Lara O'Donnell, Chris Stark, Andrew Chatr-Aryamontri, Kara Dolinski and Mike Tyers

116. **GEOmAtik: Automated platform for mining and classification of individual datasets of NCBI GEO.** Madhura Vipra and Devaki Kelkar

119. **Metabolic pathway extraction from text.** Cecile Pereira and Ana Conesa

123. **A new and integrative curation system for RegulonDB.** Socorro Gama-Castro, Fabio Rinaldi, Hilda Solano-Lira, Luis José Muñiz-Rascado, Oscar Lithgow, Cecilia Ishida-Gutierrez, Sara Martinez-Luna, Victor Hugo Tierrafría, Carlos-Francisco Méndez-Cruz, Alejandra López-Fuentes and Julio Collado-Vides

124. **SAP – a CEDAR-based pipeline for semantic annotation of biomedical metadata.** Ravi Shankar, Marcos Martinez-Romero, Martin O'Connor, John Graybeal, Purvesh Khatri and Mark Musen

143. **Gold standard evaluation of machine and human generated annotations of biodiverse phenotypes.** Wasila Dahdul, Prashanti Manda, Hong Cui, James Balhoff, Alex Dececchi, Nizar Ibrahim, Hilmar Lapp, Paula Mabee and Todd Vision

Data Standards and Ontologies

Berg Hall, Rm A; Monday, March 27, 12-1:30 PM

11. **Exposure Science in The Comparative Toxicogenomics Database: Linking Chemical Stressors to Outcomes via an Exposure Ontology.** Cynthia Grondin, Allan Davis, Jolene Wiegers, Thomas Wiegers, Benjamin King and Carolyn Mattingly

34. **Trimmed Graph Visualization of Ontology-based Annotations.** Raymond Lee, Juancarlos Chan, Christian A. Grove and Paul W. Sternberg

125. **Chinese Human Phenotype Ontology — the Chinese Semantic Standard for Phenotype.** Xiaolin Yang, Yiming Zhou, Liu Yang, Sheng Yang, Heng Wang, Bing Liu, Zhi Zhang and Jian Guan

126. **A Searchable Catalogue of Validated Antibodies Used in the ENCODE Project.** Esther Chan, Jason Hilton, Kathrina Onate, Idan Gabdank, Marcus Ho, Aditi Narayanan, J Seth Strattan, Ulugbek Baymuradov, Forrest Tanaka, Christopher Thomas, Cricket A. Sloan, Benjamin Hitz and Mike Cherry

127. **Towards the standardization of biomedical terminologies in China: from CMeSH to CMLS.** Junlian Li, Xiaoying Li, Yujing Ji, Sizhu Wu, Lin Yang and Qing Qian

132. **Curation and ontology resources used in the gene expression database Bgee.** Anne Niknejad, Amina Echchiki, Angelique Escoriza, Julien Roux, Marc Robinson-Rechavi and Frederic B. Bastian

133. **Phenotype curation in Xenbase.** Malcolm Fisher, Joshua Fortriede, Christina James-Zorn, Troy Pells, Kevin Burns, Virgilio Ponferrada, Erik Segerdell, Kamran Karimi, Praneet Chaturvedi, Vaneet Lotay, Vicente Pader, Stanley Chu, Ying Wang, Dong Zhuo Wang, Peter Vize and Aaron Zorn

135. **Development of Avian Anatomy Ontology Annotation.** Jinhui Zhang and Fiona McCarthy

140. **Methods for Ensuring Consistency and Accuracy in Data Submission for Data Coordination Centers.** Aditi Narayanan, Cricket A. Sloan, Esther T. Chan, Idan Gabdank, Jason A. Hilton, Marcus Ho, Kathrina C. Onate, J. Seth Strattan, Tim Dreszer, Ulugbek Baymuradov, Forrest Tanaka, Christopher Thomas, Benjamin Hitz and J. Michael Cherry

144. **Refactoring the Evidence & Conclusion Ontology by harmonizing with the Ontology for Biomedical Investigations.** Rebecca C Tauber and Marcus C Chibucos Phd.

Curation Standards and Best Practice, Challenges in Biocuration, Biocuration Tutorial

Berg Hall, Rm A; Monday, March 27, 12-1:30 PM

4. **Biocuration of Experimentally-Determined 3D Macromolecular Structures and their Complexes at the wwPDB**. Jasmine Young, John Berrisford, Reiko Igarashi, Wwpdb Biocuration Team, Wwpdb Onedep Team, John Markley, Haruki Nakamura, Sameer Velankar and Stephen Burley

12. **Training Future Biocurators Through Data Science Trainings and Open Educational Resources.** Nicole Vasilevsky, Ted Laderas, Jackie Wirz, Bjorn Pederson, David Dorr, William Hersh, Shannon McWeeney and Melissa Haendel

13. **A Need for Better Data Sharing Policies: A Review of Data Sharing Policies in Biomedical Journals.** Nicole Vasilevsky, Jessica Minnier, Melissa Haendel and Robin Champieux

32. **Introducing the Tag Storm format.** Clayton Fischer

46. **Training needs for biocuration workshop report**. Claire O'Donovan, Sangya Pundir, Marc Robinson-Rechavi and Patricia Palagi

75. **Using Shape Expressions to model, validate and curate Wikidata.** Andra Waagmeester, Eric Prud'Hommeaux, Elvira Mitraka, Gregory Stupp, Núria Queralt-Rosinach, Sebastian Burgstaller-Muehlbacher, Timothy Putman, Benjamin Good and Andrew I. Su

80. **Biocuration as an undergraduate training experience: Improving the annotation of the insect vector of Citrus greening disease.** Surya Saha, Prashant Hosmani, Krystal Villalobos-Ayala, Sherry Miller, Teresa Shippy, Andrew Rosendale, International Psyllid Sequenciong And Annotation Consortium, Xiaolong Cao, Haobo Jiang, Chris Childers, Mei-Ju Chen, Mirella Flores, Wayne Hunter, Michelle Cilia, Lukas Mueller, Monica Munoz-Torres, David Nelson, Monica Poelchau, Josh Benoit, Helen Wiersma-Koch, Tom D'Elia and Susan Brown

96. **GOBLET Standards Committee: best practices in standards in bioinformatics and biocuration.** Maria Victoria Schneider

101. **Best Practices for Data Provenance in Wikidata**. Gregory Stupp, Timothy Putman, Sebastian Burgstaller-Muehlbacher, Andra Waagmeester, Andrew Su, Benjamin Good and Núria Queralt-Rosinach

121. **GrainGenes Update: Curating New Resources For the Small Grains Community.** Sarah G. Odell, Gerard R. Lazo, David L. Hane, Yong Q. Gu and Taner Z. Sen

142. **Ameliorated, exacerbated, and biomarker phenotype annotation at ZFIN.** Yvonne Bradford, David Fashena, Ceri Van Slyke, Christian Pich and Zfin Staff

152. **The Drug Repurposing Library: Curating a collection of clinical compounds for novel therapeutic discovery.** Zihan Liu, Jodi Hirschman, Joshua Gould, Joshua Bittker, Patrick McCarren, Bang Wong, Mariya Khan, Jacob Asiedu, Aravind Subramanian, Todd Golub and Steven Corsello

---

Curation for Precision Medicine

Berg Hall, Rm A; Monday, March 27, 12-1:30 PM

35. **Curation of human protein variants in UniProtKB/Swiss-Prot.** Lionel Breuza and Uniprot Consortium

42. **Using literature to predict relevant mutations for cancer treatment**. Emilie Pasche, Anaïs Mottaz, Franziska Singer, Nora Toussaint, Daniel Stekhoven and Patrick Ruch

60. **hgvs-eval: automated evaluation suite to access HGVS-formatting tools.** Nicole Ruiz-Schultz, Justin Paschall, Xing Xu, David Caplan, Carolyn Ch'Ng, Karen Eilbeck and Reece Hart

78. **Potentials of databases of biocomputational models for precision medicine.** Esra Bas

81. **Variant Coordinate Curation for Variant Knowledgebases, the CIViC approach.** Kilannin Krysiak, Nicholas Spies, Lynzey Kujan, Cody Ramirez, Benjamin Ainscough, Adam Coffman, Joshua McMichael, Arpad Danos, Erica Barnell, Alex Wagner, Connor Liu, Zachary Skidmore, Yan-Yang Feng, Katie Campbell, Elaine Mardis, Obi Griffith and Malachi Griffith

105. **The BioGRID Interaction Database: Curation and Network Visualization of Genetic, Protein and Chemical Interactions for Drug Discovery and Drug Repurposing.** Rose Oughtred, Bobby-Joe Breitkreutz, Lorrie Boucher, Christie Chang, Jennifer Rust, Andrew Chatr-Aryamontri, Nadine Kolas, Lara O'donnell, Chandra Theesfeld, Chris Stark, Kara Dolinski and Mike Tyers

107. **Whole-genome reference panel of Tohoku Medical Megabank Organization (ToMMo) and biomedical variant annotation for estimating frequencies of pathological variants in the Japanese population.** Yumi Yamaguchi-Kabata, Yosuke Kawai, Kaname Kojima, Takahiro Mimori, Fumiki Katsuoka, Shigeo Kure, Yoichi Suzuki, Nobuo Fuse, Hiroshi Kawame, Masao Nagasaki, Jun Yasuda, Kengo Kinoshita and Masayuki Yamamoto

109. **COSMIC: expanding curation to highlight drug-resistant mutations in cancer.** Laura Ponting, Sally Bamford, Charlotte Cole, Sari Ward, Elisabeth Dawson, Raymund Stefancsik, Nidhi Bindal, David Beare, Harry Boutselakis, Bhavana Harsha, Mingming Jia, Harry Jubb, Chai Yin Kok, Claire Rye, Zbyslaw Sondka, John Tate, Sam Thompson, Shicai Wang, Simon Forbes and Peter Campbell

114. **Exploring the link between NSAIDs and variable cardiovascular risk response in the literature: The PENTACON Curated Data Resource (CDR) suite.** Jennifer Rust, Rose Oughtred, Michael Livstone, Christie Chang,

Katie Theken, Faith Coldren, Chandra Theesfeld, Jodi Hirschman, Alicja Tadych, Sven Heinicke, John Matese, Robert Murphy, Tilo Grosser, Garret Fitzgerald, Olga Troyanska, Anastasia Baryshnikova and Kara Dolinski

115. **IMGT® biocuration of IG and TR in IMGT/LIGM-DB and IMGT/GENE-DB.** Joumana Jabado-Michaloud, Géraldine Folch, Marie-Paule Lefranc, Véronique Giudicelli, Patrice Duroux, Sofia Kossida, Safa Aouinti, Mélissa Cambon, Imène Chentli, Saida Hadi-Saljoqi, Karthik Kalyan, Anjana Kushwaha, Arthur Lavoie, Claudio Lorenzi, Perrine Pégorier and Laurène Picandet

146. **Data Curation at cBioPortal.** Ritika Kundra, Hsiao-Wei Chen, Adam Abeshouse, Debyani Chakravarty, Ino de Bruijn, Jianjiong Gao, Benjamin Gross, Zachary Heins, Moriah Nissan, Angelica Ochoa, Sarah Phillips, Julia Rudolph, Robert Sheridan, Onur Sumer, Yichao Sun, Jiaojiao Wang, Manda Wilson, Hongxin Zhang and Nikolaus Schultz

147. **ClinGen's Gene and Variant Curation Interface Suite: Centralized and Consistent Evaluation of the Clinical Relevance of Genes and Variants.** Matt W. Wright, Selina Dwight, Karen Dalton, Minyoung Choi, Jimmy Zhen, J. Michael Cherry and Clinical Genome Resource (ClinGen)

151. **Creation of biomedical concept dictionaries for applications in rare disease gene prioritization.** Aditya Rao, Thomas Joseph, Sujatha Kotte, Saipradeep Vangala, Prisni Rath, Naveen Sivadasan and Rajgopal Srinivasan

Workshops

Workshop 1: GigaScience Curation Challenge

Organizers: Chris Hunter, Todd Taylor, Maryann Martone

Summary: This workshop will introduce community annotation tools, iCLiKVAL and Hypothes.is and challenge curators to use these tools. There will be three short presentations:

1. iCLiKVAL introduction and use of - by an iCLiKVAL team member
2. Hypothes.is introduction and use of - by an hypothes.is team member
3. Competition outline, rules and registration details - by Chris Hunter (GigaScience)

If time allows, there will be a short hands-on trial/mini competition session.

Time: Sunday March 26, 2017, 1:30-3:30 PM
Location: LK 120

Workshop 2: Reading, Assembling and Reasoning for Biocuration

Organizers: Sophia Ananiadou, Riza Batista-Navarro, Paul Cohen, Diana Chung, Emek Demir, Lynette Hirschman, Parag Mallik

Summary: We will focus on recent advances in the development of integrated systems to capture "Big Mechanisms" for biological systems, including machine reading of journal articles, (semi-)automated assembly of signaling pathway models, and machine-aided analysis of these models for tasks such as drug repurposing and explaining drugs' effects. This workshop will consist of invited speakers and contributed talks and/or panel discussions from experts in biocuration, machine reading, and biological modeling.

Time: Sunday March 26, 2017, 1:30-3:30 PM
Location: Berg Hall B/C – LK 240/250

Workshop 3: Addressing the High Throughput, Low Information Data Crisis in Biology

Organizers: Sean Mooney, Predrag Radivojac, Claire O'Donovan, Iddo Friedberg

Summary: This workshop aims to improve the understanding of protein function prediction methods, database biases, and the Critical Assessment of Functional Annotation (CAFA) challenge. We will also discuss how to improve automatic annotation, reduce database bias, and increase annotation accuracy. There will be four talks, followed by a group discussion:

- Sean Mooney: Introduction to the world of community challenges.
- Predrag Radivojac: Introduction to function prediction, and CAFA
- Iddo Friedberg: Understanding annotation bias in biological databases
- Claire O'Donovan: The ECO ontology as a solution to annotation biases

Time: Sunday March 26, 2017, 1:30-3:30 PM
Location: LK 130

Workshop 4: Biocuration and the Research Life Cycle: Advances and Challenges

Organizers: Cecilia Arighi, Pascale Gaudet, Lynette Hirschman, Rezarta Islamaj-Dogan, Fabio Rinaldi

Summary: This workshop will revisit and identify the major advances and new challenges in the biocuration workflow in connection to the research cycle, from publication to data acquisition to a database entry and subsequent updates. Brief introduction to the different topics (15 min), followed by breakout sessions to discuss those topics (1h), and concomitant report from each group on the outcomes and future steps (30 min). The last 15 min will be used for general discussion and workshop closing.

Time: Tuesday, March 28, 2017, 1:30-3:30 PM
Location: Berg Hall B/C – LK 240/250

Workshop 5: Google Summer of Code

Organizers: Marc Gillespie, Robin Haw

Summary: The Open Genome Informatics group will be discussing Google Summer of Code, a fantastic platform for student training, project development, and collaboration. All of these are key aspects of a good biocuration project, and in our experience the student projects result in valuable deliverables. There will be an introduction followed by a panel discussion.

Time: Tuesday, March 28, 2017, 1:30-3:30 PM
Location: Berg Hall A – LK 230

Workshop 6: Consensus Building for Cancer Molecular Subtyping

Organizers: Lynn Schriml, Sherri De Coronado, Warren Kibbe, Pascale Gaudet, Raja Mazumder

Summary: This workshop's goal is to bring together community members to identify common and alternative methods of molecular modeling. We will be exploring the status, mechanisms, and uses for molecular characterizations of cancer, ways of defining cancer subtypes, and the relations between subtypes and associated data (e.g., anatomy, OMIM phenotype 'susceptibility_to', animal models, drug modeling).

Time: Wednesday, March 29, 2017, 8:30-10:30 AM
Location: Alway M106

Workshop 7: Scientific Evidence for Biocuration

Organizers: Marcus Chibucos

Summary: This workshop, hosted by the Evidence & Conclusion Ontology (ECO), will introduce fundamental concepts of representation of scientific evidence, give new users an overview of recent ECO developments, applications, and collaborations, serve as an open forum for discussion of community evidence needs, including confidence/quality metrics, and invite new collaborations and users. There will be an introductory talk followed by an open discussion.

Time: Wednesday, March 29, 2017, 8:30-10:30 AM
Location: Berg Hall A – LK 230