CHIARA SABATTI

Dec 1st, 2025

PERSONAL DATA

Computing and Data Science (CoDa), 389 Jane Stanford Way, Stanford, CA 94305-5464
Phone (650) 725 6783 | e-mail `sabatti@stanford.edu`
`https://chiarasabatti.su.domains`

ACADEMIC HISTORY

Degrees

B.S. and M.S. summa cum laude in Economics and Statistics, Bocconi University, Milan, Italy (1993).
Professor Eugenio Regazzini, advisor.

Ph.D. in Statistics, Stanford University, Stanford, CA (1998).
Professor Jun Liu, advisor.

Postdoctoral Training

Genetics Department, Stanford Medical School, Stanford, CA (1998-2000).
Professor Neil Risch, advisor.

Professional Training

Principles and Practices of Effective Teaching (2019), taught by Prof. Wieman.

Stanford Medicine Leadership Academy, cohort 4 (2020-21).

EMPLOYMENT

Current Professor of Biomedical Data Science and Statistics, Stanford University. Member of BioX and of the Cancer Center. Associate Director of the interdisciplinary major in Data Science (September 2022–present) and Mathematical and Computational Science (September 2019–August 2025). Associate Chair of Education and Training in Biomedical Data Science (May 2020–present). Associate Director for Education of Stanford Data Science (September 2020–present).

September 2018–February 2019 Vice Chair of Biomedical Data Science.

November 2015–July 2016 Associate Professor of Biomedical Data Science, Stanford University.

January 2015–July 2016 Associate Professor of Statistics, Stanford University.

September 2009–October 2015 Associate Professor of Health Research and Policy, division of Biostatistics, Stanford University.

July 2009–August 2011 Professor of Human Genetics and Statistics, University of California, Los Angeles. (On leave for part of the period.)

July 2006–June 2009 Associate Professor of Human Genetics and Statistics, University of California, Los Angeles. Associate of UCLA-NASA Institute for Cell Mimetic Space Exploration (2003–2008). Associate of the UCLA Center for Society and Genetics (2004–2009).

July 2000–June 2006 Assistant Professor of Human Genetics and Statistics, University of California, Los Angeles.

HONORS AND MEMBERSHIPS IN PROFESSIONAL SOCIETIES

- Fellow of the Institute of Mathematical Statistics (2022).
- "Vasa cube" award for continuous commitment to *Fiat lux* seminars (2008).
- NSF Division of Mathematical Sciences CAREER award (2003).
- UCLA Career award (2002-3)
- Statistics Department Teaching Assistant Award, Stanford University (1998)
- Doctoral Research Fellowship of the Italian Ministry of University and Scientific Research, declined (1993)
- Fellowship "Amici della Bocconi" (1993-1994)
- Fellowship "Credito Bergamasco" (1993 and 1991)

Member of the American Society of Human Genetics, the American Statistical Association, the Institute of Mathematical Statistics, and the International Society for Computational Biology.

PUBLIC AND PROFESSIONAL SERVICE

Editorial Board Membership

- Statistical Learning and Data Science (2025–)
- Statistical Science (2020–2022)
- Harvard Data Science Review (2019–present)
- Genetics (2012–2024)
- Journal of the American Statistical Association (2011–2015)
- The Annals of Applied Statistics (2010–2022)
- BMC Bioinformatics (2010–present)
- IEEE/ACM Transactions on Computational Biology and Bioinformatics (2004-2010)
- Biology Direct (Genomics, Bioinformatics, and Systems Biology section)
- Member of Faculty of 1000.

Guest Editor

- "Data Science in a Time of Crisis: Lessons from the Pandemic," special issue of Statistical Science (2022). Co-guest editor with John Chambers

Journal Reviewer

Statistical Methodology The Annals of Applied Statistics, Bayesian Analysis, Biometrics, Biostatistics, IEEE Transactions on Image Processing, IEEE Transactions on Neural Networks, the Journal of the American Statistical Association, the Journal of Computational and Graphical Statistics, the Journal of the Royal Statistical Society B, the Journal of Statistical Planning and Inference, Statistica Sinica, Statistics and Computing, Statistics and Probability Letters.

Genetics and computational biology The American Journal of Epidemiology, the American Journal of Human Genetics, the Annals of Human Genetics, Bioinformatics, Biological Psychiatry, BMC Bioinformatics, BMC System Biology, the FASEB Journal, Genetic Epidemiology, Genetics, Human Heredity, the Journal of Cellular Biochemistry, Mathematical Biosciences, Nature Genetics, Nature Biotechnology, PLOS Genetics, PLOS Computational Biology, Statistical Applications in Genetics and Molecular Biology.

Science at large Proceedings of the National Academy of Science, Science.

Membership in grant review panels

- NIH NHGRI Data Integration and Statistical Analysis Methods – DISAM (March 2024)
- National Science Foundation (NSF) review panel for Statistics (March 2023)
- DMS/NIGMS review panel (December 2020)
- NIH, BMDR study section (June 2020)
- NIH, center for genomic excellence, ad hoc reviewer (November 2019)
- NIH, ad hoc review panel (October 2017).
- NIH, review panel for T32 (Biomedical Big Data) (October 2016).
- National Science Foundation (NSF) review panel for Statistics (January 2016).
- NIH Special Emphasis Panel on Genes, Genomes and Genetics (June 2015).
- NIH Special Emphasis Panel of the Therapeutic Approaches to Genetic Diseases (July 2014).
- NIH Special Emphasis Panel on Genetics of Health and Disease (November 2013).
- NIH Special Emphasis Panel on Computational Genetics and Genomics (March 2011).
- Genomics, Computational Biology and Technology study section [GCAT] (guest member February 2008, October 2008, October 2010).
- NIH Biostatistical Methods and Research Design study section (guest member, October 2007).
- National Science Foundation (NSF) review panel for Statistics and Probability (2003).

Ad hoc grant reviews

Reviewer of grant proposals to the National Science Foundation (DMS 2005, 2003; and MCB 2009, 2004), the Welcome Trust (2005, 2006), the UCLA AIDS institute, and the Netherlands Organization for Scientific Research, the United States–Israel Binational Science Foundation.
Reviewer of proposals for the Spectrum Pilot grants "Population Health Sciences and Community Engagement" (2012), the Seed grant opportunity in Big Data for Human Health (Winter 2014), the Stanford Data Science Initiative (Fall 2014).

External review of organizations

- Chair of the 2025 NSF external review panel for the National Institute for Theory and Mathematics in Biology.
- Assessment of the Centre National de la Recherche Scientifique (CNRS) in 2023.
- Program Review of the Department of Biostatistics at the University of North Carolina at Chapel Hill (2019)

Organization of Scientific Conferences

- "Weaving Data Science into the Fabric of the University," Stanford Data Science Inaugural Conference, Stanford, April 5th, 2022.
- "Data Practices," a workshop hosted by Stanford Data Science, CESTA (Center for Spatial and Textual Analysis) and Stanford Libraries, October 2020
- IMS sponsored section at the Joint Statistical Meetings 2011: "Bioinformatics and Genomics," Miami, July 31st, 2011.
- Workshop "Computational Genetics," February 9, 2007, IPAM, Los Angeles.
- Session on "Statistical Genomics" for Interface 2006 "Massive Datasets and Streams," Pasadena, May 24-27, 2006.
- Workshop: "Sequence analysis towards system biology," IPAM January 9-13, 2006. (Co-organizer with Jun Liu.)
- Invited session titled "Genome-wide Association Analysis: Finding the Needles in the Haystack," at the ASHG Annual Meeting, Salt Lake City, October 25-29, 2005. (Co-organizer with M. Ritchie.)
- Member of the International Program Committee of the 27th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, September 1-5, 2005, Shanghai, China.
- Chair of the session on "Computational Genomics and Proteomics," 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, September 1-5, 2004, San Francisco.

Service in Professional Societies

Chair of the IMS Committee on Special Lectures (2023-24; member in 2022-23, 2024-25).

Outreach

Founding organizer of "Data science mentoring" (BIODS360), a program which strives to recruit students from backgrounds that are underrepresented in data science, reaching them early enough to make an impact on their undergraduate education and equip them for successful graduate studies. Stanford doctoral students (mentors) and undergraduate students from non-R1 universities (participants) meet weekly one-on-one using an on line platform. In the course of the class, the Stanford students are trained in mentoring and teaching, with an intentional commitment to inclusion. The participants are exposed to work and research in data science; receive coaching in planning courses of study and advice on navigating internship opportunities and preparing applications; get tutoring in aspects of data science; and engage in mini-research projects. `https://dbds.stanford.edu/data-science-mentoring/`
Founding organizer of Stanford Data Science Undergraduate Research Pathways (DSURP). The Stanford Data Science Undergraduate Research Pathways program is an 8-week full-time research experience designed to provide students at institutions without access to research opportunities the chance to conduct a research project under the supervision of both a mentor and faculty member.n `https://datascience.stanford.edu/education/undergraduate-research`

Committee Service

Stanford

- Chair Academic Council Committee on Undergraduate Standards and Policy (C-USP) 2024–2025.
- Chair of Search Committee for Faculty in Biomedical Data Science 2023-24.

- Chair of the Committee on Diversity, Equity, Inclusion and Belonging for the Statistics department (AY: 2023–2024).
- Member of the Academic Council Committee on Undergraduate Standards and Policy (C-USP) 2022–2024.
- Committee on an interdepartmental Master in Data Science (Winter 2020–2021)
- Committee on the creation of a new Major in Data Science (Fall 2019–May 2021)
- Faculty search committee for HRP Biostat (fall-winter 2015-16), Biomedical Data Science (fall-winter 2015-16), and Statistics (fall-winter 2015-16; 2017-8, and fall-winter 2024-25); psychiatry (sleep; 2018); Data Science and Neuroscience (fall-winter 2021-22; fall-winter 2022-23).
- Faculty search committee for Associate Director of Population Science (Stanford Cancer Center, 2018)
- Member of the Data Science Design Team (Fall 2018–present)
- Member of the natural sciences pathways committee (2018)
- Co-chair of the committee for Planning for Undergraduate education in the Physical Sciences (winter 2017);
- Steering committee for Biomedical Data Science (summer and fall 2014; summer 2015);
- Executive committee for the France-Stanford Center for Interdisciplinary Studies (2012–present);
- Executive committee for the Center for Computational, Evolutionary, and Human Genomics (2012–present);
- Big Data in medicine initiative (2013–2014);
- Judicial Panels (2011– 2013);
- Strategic planning group for the Children's Health Research Institute at Stanford (2012);
- Search committee for an Associate Director for Population Sciences in the Stanford Cancer Center (2011-12);
- Population Sciences Focus Group (2011).

UCLA

- Provost's committee on faculty work/life issues (2008–2009)
- Executive committee for the Interdepartmental Program in Bioinformatics (2007–2009).
- Executive Committee on Information Technology for the Life Sciences in the UC-Industry Cooperative Research Program (2005–2009).
- UCLA Early Care and Education advisory board (2003–2009).
- Childcare Subcommittee of the Academic senate faculty welfare committee (2006–2007).

RESEARCH GRANTS

Research projects for which C.S. is the principal investigator are indicated with bold fonts.

R90HD118650 (PI: Simons)
NIH 7/1/24–6/30/29
Interdisciplinary research training in maternal and childhood pain

1R01AG089509-01A1 (PI: He)
NIH/NIA 5/1/25–3/30/30
Interpretable machine learning methods for the analysis of Alzheimers disease genetics

UM1TR004921(PI: O'Hara)
NIH 7/1/24–6/30/31
Stanford Center for Clinical & Translational Education and Research

DMS 2210392 (PI: Sabatti)
National Science Foundation 9/1/22–8/31/25
Scientific Findings across Multiple Environments: Replication, Robustness, and Equity in Genetic Association Studies

R56HG010812 (PI: Sabatti)
National Institutes of Health/NHGRI 1/1/21–12/31/23
The pursuit of genetic causal mechanisms

R01MH123157 (Freimer, Stanford PI: Sabatti)
UCLA/ NIH primary 5/1/20–2/28/26
A Latin American biobank for large-scale genetics research on severe mental illness

1934578 (PI: Candès, Micheli, Sabatti, Leskovec)
NSF 9/1/19–8/31/21
The Stanford Data Science Collaboratory

R01MH113078 (PI: Freimer; Stanford PI: Sabatti)
NIH /NIMH primary, subcontract from UCLA 4/1/17–3/31/22
Genetics of Severe Mental Illness

Discovery Science Innovation Fund (PI: Sabatti)
Stanford School of Medicine 10/1/17–9/30/20
Reproducible identification of cancer cell types via scRNAseq

DMS 1712800 (PI: Sabatti)
NSF 9/01/17–8/31/20
Discovering what matters: informative and reproducible variable selection with applications to genomics

01MH110437 (PI: Zandi; Stanford PI: Sabatti)
NIH primary, subcontract from John Hopkins University 8/01/16–5/31/19
The Bipolar Sequencing Consortium for Combined Analysis and Follow-Up

R01 MH101782 (PI: Sabatti)
NIH/NIMH 8/1/13–7/31/18
Genetic Regulation of Gene Expression and its Impact on Phenotypes

4UL1 TR001085 (PI: Greenberg)
NIH 9/26/13–4/30/18
Stanford Center for Clinical and Translational Education and Research

Discovery Science Innovation Fund (PI: Olshen, Sabatti, Tian)
Stanford School of Medicine 9/1/14–8/31/15
How do genes influence the interactions between multiple traits?

Discovery Science Innovation Fund (PI: Olshen, Sabatti, Tian)
Stanford School of Medicine 9/1/13–8/31/14
New Approaches to Quantifying the Phenotype-Genotype Relationship

R01 HL118612 (PI: Nadeau and Miller)
NIH 9/1/13–8/31/17
Secondhand Smoke and Asthma: Mechanistic Outcomes of DNA Methylation in T Cells

U01 HG007419 (PI: Matise; Stanford PI: Bustamante)
NIH /NHGRI, subcontract from Rutgers 6/1/13–5/31/17
NHGRI PAGE Coordinating Center

R01 HG006695 (PI: Sabatti)
NIH/NHGRI 4/1/13–3/31/17
New Statistical Methods for High Resolution Mapping of Multiple Phenotypes

R21 CA185804 (PI: Wang)
NIH 12/2/14–11/30/16
Analysis of CTC's for Early Prediction of Response to Treatment in RCC

R21 CA169964 (PI: Felsher)
NIH/NCI 8/1/12–7/31/14
Nanoscale proteomic profiles of hypoxia pathways to develop biomarkers of renal cell carcinoma

T32 GM096982 (PI: Tibshirani, Olshen, Sabatti)
NIH/NIGMS 7/1/12–6/30/17
Interdisciplinary Training Program in Biostatistics

R01 HL113315 (PI: Freimer, Palotie; Stanford PI: Sabatti)
NIH/NHLBI, subcontract from UCLA 6/15/12–3/31/17
Genomic and Metabolomic Profiling of Finnish Familial Dyslipidemia Families

R01 ES02092601 (PI: Nadeau)
NIH 6/1/12–2/28/17
Phenotyping/Epigenetic studies of PAH-associated Treg impairment in asthma

R01 MH095454 (PI: Freimer; Stanford PI: Sabatti)
NIH/NIMH, subcontract from UCLA 6/1/12–5/31/15
Genome Sequencing in Extended Bipolar Pedigrees

RC2 AG036607 (PI: Schaefer, Risch; Stanford PI: Tang)
NIH (ARRA), subcontract from Kaiser 9/30/09–8/31/12
A resource for genetic epidemiology research in adult health and aging

R01 GM053275-14 (PI: Lange)
NIH/NIGMS 4/1/08–3/31/13
Statistical Methods for Gene Mapping

P30 1MH083268 (PI: Bilder)
NIH 9/1/07–8/31/12
Interdisciplinary Research Consortia – Whole Genome Association Analysis Strategies for Multiple Phenotypes

R01 MH075007 (PI: Freimer)
NIH/NIMH 12/1/06–5/30/11
Bipolar Endophenotypes in Population Isolates

R01 HL087679-01 (PI: Peltonen)
NIH 10/10/06–9/30/09
Genetics of Cardiovascular Risk Factors in a Large Birth Cohor from a Founder Population

R01 MG078075-01 (PI: Ophoff)
NIH/NIMH 7/1/06–6/30/10
Genome-wide Association Study of Schizophrenia

R01 NS37484 (PI: Freimer)
NIH 4/1/04–3/31/09
Population Genetic Mapping of Tourette Syndrome

R01 GM53275 (PI: Lange)
NIH/NIGMS 4/1/04–3/31/08
Statistical Methods for Gene Mapping

CCF-0326606 (PI: Liao)
NSF 12/3/03-12/2/07
ITR:"Regulography"- Quantitative Reconstruction of Transcriptional Regulatory Networks

DMS0239427 (PI: Sabatti)
NSF 6/1/03–5/31/08
CAREER: Statistical and Computational Tools for the Analysis of High Dimensional Genetic Data

NCC2-1364 (PI: Ho)
NASA/AMES research center 9/1/02–8/30/07
Institute for Cell Mimetic Space Exploration (CIMSE)

TEACHING

Stanford Courses

BIODS 232/BMDS 291 Consulting Workshop on Biomedical Data Science (aka “Data Studio”), continuously since Fall 2018.

BIODS 240 Race, Data, Algorithms and Health, Fall 2020. I introduced this class which partially fulfills the "Social and Ethical Issues" BMI requirement. It explores how data science interplays with the use and misuse of racial information in the study of biology and the delivery of healthcare.

BIODS 360 Data Science Mentoring, Spring 2021, Winter 2022, Winter 2023, Winter 2024, Winter 2025. I introduced this class which connects outreach and diversity efforts with the fundamental goals of educational institutions. Stanford students learn about inclusivity while mentoring an undergraduate student attending a non research-oriented university.

DATASCI 120 Data Narratives, Spring 2022, Spring 2023, Spring 2024, Spring 2025. I designed this class to fulfill the “writing in the major” requirements for the new Data Science major.

DATASCI 190 The Data Science Experience, Spring 2024, Spring 2025. This is one class that I designed as part of the Data Science BS Capstone options (I have been in charge in the past two years to design this entire set of experiences for the new major).

Stats 48N: Riding the data wave, Fall 2012, Fall 2013, Fall 2014, Fall 2015, Fall 2016, Fall 2019, Fall 2020. I designed this introductory seminar to provide students with the awareness and some of the tools necessary to be citizens in our data rich world. `https://chiarasabatti.su.domains/Stat48/`

Stats 101: Data Science 101, Spring 2017, Spring 2018. I was a member of the “Data Science Working group” that designed this new class.

Stats 200: Introduction to Statistical Inference, Winter 2018, Winter 2021.
`https://chiarasabatti.su.domains/Stat200/`

Stats 319: Statistic Literature: Inference and Selection, Winter 2014; Statistic Literature: Algorithmic Fairness, Spring 2020;

Stats 367: Statistical Models in Genetics, Fall 2011, Winter 2015
`https://chiarasabatti.su.domains/Stat367/`

Stats 370: A course in Bayesian Statistics, Winter 2013, Winter and Spring 2016.
`https://chiarasabatti.su.domains/Stat370/`

Stats 390: Consulting Workshop, Fall 2011.

Summer Fellowships

Faculty in charge of the Data Science for Social Good summer program at Stanford, Summers 2019, 2020.
`https://datascience.stanford.edu/programs/data-science-social-good-summer-program`.

Faculty in charge of the SURP-Stats for summer 2023, 2024, 2025.

Founding faculty for Stanford Data Science Undergraduate Research Pathways (DSURP) 2024.

UCLA Courses

Stat 202c: Monte Carlo Methods in Statistics. (2007-09)

Hum Gen 236: Advanced human genetics (2001-09) Responsible for lectures on population genetics and linkage disequilibrium.

Stat 180–236: Introduction to Bayesian statistics (2001-08).

HG 19–sem 2: Genetics enhancements and the quest for perfection. A fiat lux seminar, Winter 2008, Spring 2009.

Stat 19–sem 2: The science and morality of Genetics. A fiat lux seminar, Fall 2006.

Stat 254: Statistical methods for computational biology. (2002-2005) A core course for the interdepartmental PhD program in Bioinformatics.

Stat 189: Honors section for Stat 180, Winter 2005: “Early progress in statistics and links to the eugenics movement.”

Stat 19–sem 5: Eugenics and statistics. A fiat lux seminar, Fall 2005.

Stat 19–sem 4: Decoding bioinformation. A fiat lux seminar, Fall 2003.

Hum Gen 210: Medical genetics. UCLA, Fall 2000. Responsible for a lecture on linkage and linkage disequilibrium methods.

Short Courses

- “Reproducibility in data science: some statistical tools and applications,” Bocconi Summer School in Advanced Statistics and Probability, July 6-17 2020 (cancelled due to the COVID-19 pandemic).
- “Workshop on Statistical Genetics,” Monterrey, Mexico, September 16-19 2019.
- “XII CIMAT School in a Probability and Statistics,” CIMAT in Guanajuato, Mexico, March 2014.
- “8th Annual Statistical Genetics Short Course for Obesity & Nutrition Researchers,” San Diego, April 2008.
- “Statistical Genetics Short Course, featuring Mendel,” UCLA, August 2006, August 2004, August 2001

- “Statistical genetics with Mendel,” University of Pavia, July 2005.
- “Linkage and association analyses using mendel and SimWalk software,” IGES meeting, November 2003.
- “DNA microarray workshop,” UCLA September 14-15, 2002.
- “Bayesian genomics,” Università degli Studi di Pavia, June 2002.
- “DNA Microarrays: The New Frontier in Gene Discovery and Gene Expression,” Neuroscience 31st annual meeting in San Diego, November 10, 2001.
- “DNA microarray: principles and biotechnological applications,” UCLA September 10-13, 2001.

Student Advising – Stanford

Undergraduate pre-major advisor for a total of 13 Freshmen (2010-12).

Undergraduate advisor for the inter-departmental major in Mathematical and Computational Science (2012–present), Data Science (2022–present), and Biomedical Computation (2021–present) (29 current advisees)

Master students Provided research opportunities for Alden Timme (MS in Statistics, 2012) and Hadi Zarkoob (MS in MS&E, 2012).

Doctoral dissertation committee Yash Nair (Statistics); Ying Yi (Statistics, Ph.D. 2024); Zhaomeng Chen (Statistics, Ph.D. 2025); Julie Zhang (Statistics, Ph.D. 2025); Abel Peirson (Physics, Ph.D. 2023); Kelley Paskov (Biomedical Informatics, Ph.D. 2023); Shuangning Li (Statistics, Ph.D. 2022); Julia Olivieri (ICME, Ph.D. 2022); Nicole Ferraro (Biomedical Informatics, Ph.D 2021); Zhimei Ren (Statistics, Ph.D. 2021); Stephen Bates (Statistics, Ph.D. 2020); Matteo Sesia (Statistics, Ph.D. 2020); Jesse Min Zhang (EE, Ph.D. 2019); Dangna Li (ICME, Ph.D. 2018); Zhou Fan (Statistics, Ph.D. 2018); Snigdha Panigrahi (Statistics, Ph.D. 2018); Joey Arthur (Statistics, Ph.D. 2018); Henry Li (Structural Biology, Ph.D. 2017); Jingshu Wang (Statistics, Ph.D. 2016); Olivia Liao (Statistics, Ph.D. 2013); Arwen Meister (ICME, Ph.D. 2013); Gourab Mukherjee (Statistics, Ph.D. 2013); Bokyung Choi (Applied Physics, Ph.D 2012); Nicholas Johnson (Statistics, Ph.D. 2011); Zhengqing Ouyang (Biology, Ph.D. 2010).

Doctoral Advisor for Paula Gablenz (2022–2025); Eugene Katsevich (2016–19; now on the faculty at UPenn) and Junjie Zhu (2016–2020; now at Apple).

Postdoctoral advisor Yusuph Mavura (2025–); Qian Zhao (2022–2024, now at UMass Amherst); for Benjamin Chu (2021–2025, now at Genentech); for Christine Peterson (2014–2016, now at Rice) and for Zhongyang Zhang (2012, currently at Mount Sinai).

Training grants Co-director of the Training Program in Biostatistics for Personalized Medicine (2012-17); mentor for the training grant in Biomedical Informatics.

Educational program direction Associate director of Mathematical and Computational Science (2019–present) and of Data Science B.S. (2022–present); member of the Biomedical Informatics Executive committee (2020–2024); associate director for education of the Stanford Data Science Initiative (2020–present); member of the Biomedical Computation major executive committee (2021–present).

Student Advising – UCLA

Doctoral Advisor for the Statistics Ph.D. students Zhongyang Zhang (2008–2012, currently assistant professor at Mount Sinai); Jae Brodski (2006–2010; currently at Affymetrix); Hui Wang (2003–2006; currently at the Palo Alto VA); and Steve Erickson (2002–2006; recipient of the Chancellor fellowship and of the TAG training grant; currently at RTI international).

Master Advisor for Statistics students Chia-Ho Lin (M.S. 2005) and Kaiding Zhu (M.S. 2004).

Post-doctoral Advisor for Iouri Chepelev (September 2004–2007), currently at the Cincinnati Children's Hospital Medical Center.

Doctoral dissertation committee Alejandra Young (MBI, Ph.D. 2010); Michael Mason (Statistics, Ph.D. 2010); Samuel Strom (Human Genetics, Ph.D. 2010); Linh Tran (Chemical engineering, Ph.D. 2010); Sara Tajyar (MIMG, Ph.D. 2008); Allen Day (Human Genetics, Ph.D. 2008); Yuhan Lee (Genetics, Ph.D. 2008); Sara Tajyar (MIMG, Ph.D. 2008); Kristin Ayers (Biomathematics, Ph.D. 2008); Mark Brynilnsen (Chemical Engineering, Ph.D. 2008); Wei Sun (Statistics, Ph.D. 2007); Robert Riley (Human Genetics, Ph.D. 2007); Rebecca Krupp (MIMG, Ph.D. 2006); Ching-Ti Liu (Statistics, Ph.D. 2006); Angela Presson (Statistics, Ph.D. 2006); Ben Redelings (Biomathematics, Ph.D. 2006); Katherine Comanor (Electrical Engineering, Ph.D. 2005); Joseph Dougherty (Neuroscience, Ph.D. 2005); Yumao Lu (Electrical Engineering, Ph.D. 2005); Tianwei Yu (Statistics, Ph.D. 2005); Robert Granat (Electrical Engineering, Ph.D. 2004); Lars Rohlin (Chemical Engineering, Ph.D. 2004); Riccardo Boscolo (Electrical Engineering, Ph.D. 2003).

INVITED LECTURES

Conferences

- May 21–23, 2025, STATGEN 2025: Conference on Statistics in Genomics and Genetics: "Identifying locally important variables on a genome scale."
- December 16–19, 2024, International Conference on Statistics and Data Science, Nice (FR): "Identifying locally important variables on a genome scale."
- November 18–23, 2024, Causal Inference in Genetics: Latest developments and New Directions, Fondations des Treilles, Tourtour (FR): "Local knockoff filter: exploring how genes and environment interact."
- July 22-26, 2024, 7th International Conference of Quantitative Genetics, Vienna: "Searching for causal variants in polygenic traits."
- December 13–16, 2022, International Conference on Statistics and Data Science, Florence, Italy: "Machine Learning and Genetics."
- October 24, 2022, Lange Symposium on Statistical Genetics, UCLA: "Genomic Data Science: Some Examples of New Directions."
- March 7, 2022, WiDS: "Replication, Robustness and Interpretability: Improving How We Communicate Scientific Findings"
- February 15, 2022, *Learning from Interventions* Simon Institute, Berkeley: "Searching For Causal Genetic Mechanisms Across Human Populations."
- June 16, 2021, WNAR: "Knockoff genotypes: value in counterfeit."
- March 14–17, 2021, ENAR: "Knockoff genotypes: value in counterfeit."
- June 15-19, 2020, Luminy conference on Mathematical Methods of Modern Statistics 2: "Knockoff genotypes: value in counterfeit."
- April 29–May 1, 2019, BFF6, Duke University: "Selecting Important Features in Presence of Correlation: a Story from Genetics."

- December 17–20, 2018, Joint Statistical Event, Jerusalem: “Controlling FDR while highlighting selected discoveries.”
- October 29–November 2, 2018, Robust and High-Dimensional Statistics, Simons Institute, Berkeley: “Controlling FDR while highlighting selected discoveries.”
- October 1–5, Bertinoro Computational Biology 2018, Bertinoro, Italy: “Controlling FDR while highlighting selected discoveries.”
- July 24–24, 2018, Workshop on Principles of Adaptive Data Analysis, Simons Institute, Berkeley: “Controlling FDR while sorting through discoveries.”
- January 10–12, 2018, Algorithmic Challenges in Protecting Privacy for Biomedical Data, IPAM, Los Angeles: “Knockoff genotypes: value in counterfeit.”
- JSM 2017, July 29–August 3, 2017, Baltimore: “Selective inference in Genomics.”
- Mathematical Methods of Modern Statistics, July 10-14, 2017, Marseille: “False discovery rate control in genetics.”
- 2nd Probabilistic Modeling in Genomics Conference, September 12-14, 2016, Oxford, England: “Selective inference for gene mapping."
- Advances in Statistics, Probability and Mathematical Physics, June 10-11, 2016, Pavia, Italy: “Selective inference in genetics.”
- Joint Statistical Meetings, August 9, 2015, Seattle: “Identifying genetic variants that regulate gene expression: how to insure reproducibility,” (scheduled, but cancelled due to illness).
- Bertinoro Computational Biology—Statistical and Computational Genetics, Bertinoro, Italy, September 27-October 2, 2014: “Global error control in multiple phenotype mapping.”
- Italian Society of Statistics Meeting, Cagliari, Italy, June 11-13, 2014: “In the mist of the data deluge, how to let the interesting findings surface? Tales from genetics.” (Plenary speaker)
- 7th Annual Bayesian Biostatistics and Bioinformatics Conference, Houston, TX, February 12-14, 2014: “Multiple testing, family and selection in gene mapping.”
- BigData in Biomedicine, Stanford, May 23, 2013: “Big data and reproducibility.”
- ComBio2012, Adelaide, Australia, September 23–27, 2012: “Limiting false discoveries in mapping multiple phenotypes.”
- Joint Statistical Meetings, San Diego, July 28–August 2, 2012: “Bayesian Models for rare genetic variants.”
- BIRS workshop on Challenges and Advances in High Dimensional and High Complexity Monte Carlo Computation and Theory, Banff, March 18–23, 2012: “Rare variants in genomic studies and the potential of Bayesian models.”
- IPAM workshop in Co-ancestry, Association, and Population Genomics, November 29–December 2, 2011: “Reconstructing CNV in a central american population.”
- Joint Statistical Meetings, Miami July 30–August 4, 2011: “Reconstructing DNA copy number by penalized estimation and imputation”
- JSM 2009, Washington DC, August 1-6, 2009: “Modeling population structure in genomewide association studies.”
- DIMACS Workshop on Computational Issues in Genetic Epidemiology, Rutgers University, August 21-22, 2008: “The genetics of quantitative traits: what’s new since the days of R.A. Fisher?"

- International Chinese Statistical Association Applied Statistics Symposium, Piscataway, New Jersey, June 4-7, 2008: “Reconstructing Copy Number Variations with Illumina Platform.”
- Bioinformatics, Genetics and Stochastic Computation: Bridging the Gap, Banff Centre, Alberta, Canada, July 1-6, 2007: “Volume measures for linkage disequilibrium.”
- EMBO Practical Course on SNP Genotyping and Haploblock Analysis, Helsinki, August 21-27, 2005: “High density SNPs: genotype calling and other statistical issues."
- Joint Statistical Meetings, Minneapolis, August 11, 2005: “Transcription regulation networks."
- Joint Statistical Meetings, Minneapolis, August 8, 2005: “Bayes and empirical Bayes approaches for large-scale simultaneous hypothesis testing.” (Discussion)
- Workshop on Genomewide Association Studies, Los Angeles, April 14, 2005: “Interpreting long stretches of homozygous markers in high density SNP genotyping.”
- Plant and Animal Genome XIV Conference, San Diego, January 16, 2005: “Gaussian models for high density SNP genotyping.”
- 26th Conference of Engineering in Medicine and Biology Society, San Francisco, September 1-5, 2004: “Bayesian network component analysis.”
- 2nd Markov Chain Monte Carlo Workshop, Boston, August 27-28, 2004: “MCMC in statistical genomics.”
- at IPAM Functional Genomics Reunion Conference, Los Angeles, June 1-4, 2004: “Regulatory networks in E. Coli.”
- ENAR, Pittsburgh, March 28-31, 2004: “FDR applications in genetics.”
- MSRI Workshop on the Genetics of Complex Human Diseases, Berkeley, February, 2004: “Dictionary models and gene expression regulation.”
- 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Cancun, 17-21 September, 2003: “A Vocabulon analysis of E. Coli.”
- SNP and Haplotypes Workshop, DIMACS, Rutgers University, November 22-23, 2002: “Parsing the genome in haplotype blocks.”
- First Cape Cod MCMC Workshop, September 13-14, 2002: “Multiresolution MCMC and microarrays.”
- Southern California Statistical Genetics Meeting, USC, June 6, 2002: “Ancestral haplotype reconstruction.”
- IEEE EMBS Symposyum on Biomedical Informatics, Istanbul, June 1, 2002: “A dictionary model for genome sequences.”
- ENAR/IMS meeting, Charlotte, March 25-28, 2001: “Microarrays and statistics in high dimensions.”
- SIAM Annual Meeting, Atlanta, May 12-15, 1999: “A generalization of the Gibbs sampler.”
- Multiple Comparison Procedures 2002, Bethesda, August 5-7, 2002: “Association genome screens and FDR.” (Contributed presentation)

Seminars

- Statistics Department Seminar Series, University of Chicago, November 3, 2025: “Searching for local associations while controlling the false discovery rate."

- Statistics Department Seminar Series, Stanford University, September 30, 2025: “Searching for local associations while controlling the false discovery rate."
- Biomedical data science seminar series of the University of Virginia, November 17, 2023: “Identifying Genetics Signals Adaptively and Reproducibly.”
- International Seminar on Selective Inference, November 15 , 2023: “Catch me if you can: Signal localization with knockoff e-values.”
- Data Science Applied Research and Education Seminar, University of Toronto, January 30, 2023: “Human Populations and Gene Mapping, Some Novel Directions.”
- Department of Biostatistics, Harvard University, February 17, 2022: “Searching For Causal Genetic Mechanisms Across Human Populations.”
- International Seminar on Selective Inference, October 28, 2021: “Genetic variants across human populations & our understanding of the genetic basis of traits”
- Department of Biostatistics, Columbia University, October 13, 2021: “Genetic variants across human populations & our understanding of the genetic basis of traits”
- Department of Statistics, Harvard University, September 27, 2021: “Genetic variants across human populations & our understanding of the genetic basis of traits”
- Department of Biomedical Data Science, Stanford, September 23, 2021: “Genetic variants across human populations & our understanding of the genetic basis of traits”
- Department of Statistics, University of Michigan, Ann Arbor, April 9, 2021: “Fairness and uncertainty assessment.”
- Department of Biomedical Data Science, Biostatistics Workshop, Stanford, April 16, 2020: “Fairness and uncertainty assessment.”
- Department of Biostatistics, University of North Carolina in Chapel Hill, January 9, 2020: “Knockoff genotypes, value in counterfeits.”
- Human Genetics Unit, Indian Statistical Institute, Kolkata, December 16, 2019: “Knockoff genotypes, value in counterfeits.”
- Department of Biostatistics, University of Michigan, Ann Arbor, October 3, 2019: “Knockoff genotypes, value in counterfeits.”
- Department of Medicine, Quantitative Science Unit, Stanford, October 3, 2017: “Testing hypotheses on a tree: new error rates and controlling strategies.”
- Neyman Seminar Series, UC Berkeley, March 15, 2017: “Selective Inference in Genomics.”
- Human Genetics Speaker Series, UCLA, March 6, 2017: “Controlling the rate of GWAS false discoveries.”
- Biostatistics Workshop, Stanford University, October 15, 2015: “Controlling the false discovery rate in genome wide association studies: two stories.”
- Department of Statistical Sciences, Cornell University, October 7, 2015: “Controlling the false discovery rate in genome wide association studies: two stories.”
- Department of Statistical Sciences, Università di Padova, July 1, 2015: “TreeQTL: Selective Inference for genetic regulation of gene expression.”
- UCSF Institute for Human Genetics ‘hot topics’ series, San Francisco, January 9, 2015: “Controlling the false discovery rate in genetic association studies with multiple phenotypes.”

- Department of Statistics and Operation Research, Tel Aviv University, June 30, 2014: “Penalized estimation for model selection.”
- Department of Statistical Sciences, Università di Padova, July 4, 2013: “P-values in the Big Data age."
- Departments of Biostatistics and Statistics, UC Berkeley, April 4, 2013: “Statistical Challenges in the Analysis of Resequencing Data.”
- Departments of Biostatistics, Columbia University, February 7, 2013: “Statistical Challenges in the Analysis of Resequencing Data.”
- Results for RC2 Project: A Resource for Genetic Epidemiology Research in Adult Health and Aging: Opportunities for Research in the GERA Cohort, Symposium at NIH, June 4, 2012: “Detecting CNV in a cohort of 100,000.”
- Department of Statistics, Università di Padova, July 19, 2011: “Reconstructing DNA Copy Number by Penalized Estimation and Imputation.”
- Claremont Colleges Mathematics Colloquia, April 20, 2011: “Reconstructing DNA Copy Number by Penalized Estimation and Imputation.”
- International Speaker Series, CIHR STAGE, Toronto, April 1, 2011: “Genetics in a Finnish Birth Cohort.”
- Statistics Department, University of Toronto, March 31, 2011: “Reconstructing DNA Copy Number by Penalized Estimation and Imputation.”
- UC Berkeley Statistics and Genomics Seminar, September 16, 2011: “Reconstructing DNA Copy Number by Penalized Estimation and Imputation.”
- Department of Statistics, Università di Padova, July 5, 2010: “Model selection problems in genetic association studies: a review of current practices and open problems.”
- Mathematics Department, Politecnico di Milano, June 23, 2008: “The genetics of quantitative traits: what’s new since the days of R.A. Fisher?”
- Biostatistics Department, Stanford, November 2, 2007: “Reconstructing Copy Number Variations with Illumina Platform.”
- Computer Science Department, UCLA, April 30, 2007: “Hierarchical Bayes models for SNP probes intensities.”
- Biostatistics Department, University of Wisconsin, April 13, 2007: “Hierarchical Bayes models for SNP probes intensities.”
- Biomathematics Department, UCLA, December 7th, 2006: “Volume measures for linkage disequilibrium.”
- Statistics Department, Stanford, December 5th, 2006: “Transcription regulation in E.Coli.”
- Statistics Department, UC Berkeley, October 17th, 2006: “Volume measures for linkage disequilibrium.”
- School of life sciences, EPFL, Lausanne, January 31st, 2006: “High density genotyping: challenges and opportunities.”
- Institute for Human Genetics, UCSF, November 4, 2005: “Interpreting long homozygous segments in high density genotyping data.”

- Department of Biostatistics, UCLA, November 2, 2005: “High density genotyping: challenges and opportunities.”
- Mathematics Section, EPFL, Lausanne, June 30, 2005: “Regulatory networks in E. Coli.”
- Department of Human Genetics, UCLA, June 6, 2005: “Novel genotyping technologies: challenges and opportunities for statistical analysis.”
- Department of Ecology and Environmental Biology, UCLA, June 1, 2005: “Regulatory networks: a view from genome and transcriptome.”
- Department of Statistics, UCLA, May 17, 2005: “Genetics of quantitative traits: looking forward standing on the shoulders of giants.”
- Department of Biostatistics, The University of Pittsburgh, November 17, 2004: “Regulatory networks.”
- Computer science department, Université Libre de Bruxelles, July 9, 2004: “A bioinformatics approach to transcription regulation in E. Coli.”
- Seminar für Statistik, ETH, Zürich, July 15, 2004: “A bioinformatics approach to transcription regulation in E. Coli.”
- UCLA-DOE Institute for genomics and proteomics, UCLA, April 24, 2004: “Multiple testing in microarrays?”
- Section on Statistical Genetics, Department of Biostatistics, University of Alabama at Birmingham, April 2, 2004: “Regulatory networks in E. Coli.”
- Department of Biostatistics, UCLA, March 3, 2004: “Regulatory networks in E. Coli.”
- Department of Statistics, UCLA, February 24, 2004: “Dictionary models for regulatory regions in DNA and gene expression arrays.”
- UCLA-DOE Institute for genomics and proteomics, UCLA, March 2003: “Data analysis in E. Coli using a dictionary model.”
- Department of Mathematics, University of California at San Diego, November 14, 2002: “Dictionary models for DNA sequences and gene expression.”
- Department of Statistics, University of Paris XI, July 12, 2002: “A dictionary model for genomes.”
- Department of Mathematics, Pavia University and CNR, June 25, 2002: “Minimum description length and dictionary size selection.”
- Division of Biostatistics, USC, May 2002: “False discovery rate: a paradigm for genomewise comparisons?”
- UCLA-DOE Institute for genomics and proteomics, UCLA, April 2002: “Dictionary models for DNA sequences.”
- Department of Statistics, University of Chicago, April 2002: “Genomewide motif identification using the dictionary model.”
- Department of Human Genetics, UCLA November 2000: “Homozygosity and linkage disequilibrium.”
- Department of Information and Operations Management, USC, October 2000: “Measures of dependence and volume tests.”
- Biostatistics Workshop, Stanford University Medical School, April 22, 1999: “Comparison of some measures of linkage disequilibrium.”

CHIARA SABATTI'S PUBLICATIONS

In the .pdf document, the symbol ◁ is linked to the electronic edition of the paper, whenever this is available.

Research Papers - Peer reviewed (Journals)

1. Liu, J. and **C. Sabatti** (2000) "Generalized Gibbs sampler and multigrid Monte Carlo for Bayesian computation," *Biometrika* **87**: 353–369. ◁
2. Bressman, S., **C. Sabatti**, D. Raymond, D. de Leon, C. Klein, P. L. Kramer, M. F. Brin, S. Fahn, X. Breakefield, L. J. Ozelius and N. J. Risch (2000) "The DYT1 phenotype and guidelines for diagnostic testing," *Neurology* **54**: 1746–1752. ◁
3. Liu, J., **C. Sabatti**, J. Teng, B. Keats and N. Risch (2001) "Bayesian analysis of haplotypes for linkage disequilibrium mapping," *Genome Research* **11**: 1716–24. ◁
4. Pastinen, T., I. Jaakko, M. Perola, **C. Sabatti**, P. Tainola, M.Levander, A. Syvänen and L. Peltonen (2001) "Dissecting a population genome for targeted screening of disease mutations," *Human Molecular Genetics* **10**: 2961–2972. ◁
5. **Sabatti, C.**, S. Karsten and D. Geschwind (2002) "Thresholding rules for recovering a sparse signal from microarray experiments," *Mathematical Biosciences* **176**: 17–34. ◁
6. Karsten, S., V. Van Deerlin, **C. Sabatti**, L. Gill and D. Geschwind (2002) "An evaluation of TSA signal amplification and archived fixed and frozen tissue in microarray gene expression analysis," *Nucleic Acid Research (NAR methods online)* **30**: e4. ◁
7. **Sabatti, C.** and N. Risch (2002) "Homozygosity and linkage disequilibrium," *Genetics* **160**: 1707–1719. ◁
8. **Sabatti, C.** (2002) "Measuring dependence with volume tests," *The American Statistician* **56**: 191–195. ◁
9. **Sabatti, C.**, L. Rohlin, M. Oh and J. Liao. (2002) "Co-expression pattern from DNA microarray experiments as a tool for operon prediction," *Nucleic Acid Research* **30**: 2886–2893. ◁
10. Jen, J., C. Coulin, T. Bosley, M. Salih, **C. Sabatti**, S. Nelson and R. Baloh (2002) "Familial horizontal gaze palsy with progressive scoliosis maps to chromosome 11q23-25," *Neurology* **59**: 432–435. ◁
11. **Sabatti, C.** and K. Lange (2002) "Genomewide motif identification using a dictionary model," *IEEE Proceedings* **90**: 1803–1810. ◁
12. **Sabatti, C.**, S. Service and N. Freimer (2003) "False discovery rates in linkage and association genome screens for complex disorders," *Genetics* **164**: 829–833. ◁

13. Karsten, S., L. Kudo, R. Jackson, **C. Sabatti**, H. Kornblum and D. Geschwind (2003) “Global analysis of gene expression in neural progenitors reveals specific cell-cycle and metabolic networks,” *Developmental Biology* **261**: 165–182. ◁

14. Liao, J., R. Boscolo, Y. Yang, L. Tran, **C. Sabatti** and V. Roychowdhury (2003) “Network component analysis: reconstruction of regulatory signals in biological systems,” *Proceedings of the National Academy of Science* **100**: 15522–15527. ◁

15. Kao, K., Y. Yang, R. Boscolo, **C. Sabatti**, V. Roychowdhury and J. Liao (2004) “Determination of multiple transcription regulator activities in Escherichia Coli using network component analysis,” *Proceedings of the National Academy of Science* **101**: 641–646. ◁

16. Lee, H., H. Wang, J.C. Jen, **C. Sabatti**, R.W. Baloh and S.F. Nelson (2004) “A novel mutation in KCNA1 causes episodic ataxia without myokymia,” *Human Mutation* **24**: 536 (electronic edition). ◁

17. Jen, J. C., H. Wang, H. Lee, **C. Sabatti**, R. Trent, I. Hannigan, K. Brantberg, G. M. Halmagyi, S. F. Nelson and R. W. Baloh (2004) “Suggestive linkage to chromosome 6q in families with bilateral vestibulopathy,” *Neurology* **63**: 2376–2379. ◁

18. **Sabatti, C.**, L. Rohlin, K. Lange and J. Liao (2005) “Vocabulon: a dictionary model approach for reconstruction and localization of transcription factor binding sites,” *Bioinformatics* **21**: 922–931. ◁

19. Boscolo, R., **C. Sabatti**, J. Liao and V. Roychowdhury (2005) “A generalized framework for network component analysis,” *IEEE Transaction in Computational Biology and Bioinformatics* **2**: 289–301.◁

20. Erickson, S. and **C. Sabatti** (2005) “Empirical Bayes estimation of a sparse vector of gene expression,” *Statistical Applications in Genetics and Molecular Biology* **4**: 22 (electronic edition). ◁

21. Wang, H., Y. Lee, S. Nelson and **C. Sabatti** (2005) “Inferring genomic loss and location of tumor suppressor genes from high density genotypes,” *Journal of the French Statistical Society* **146**: 153–171 (Invited Contribution).

22. Riley, R., C. Lee, **C. Sabatti** and D. Eisenberg (2005) “Inferring protein domain interactions from databases of interacting proteins,” *Genome Biology* **6**: R89 (electronic edition). ◁

23. Falcón-Pérez, J., R. Nazarian, **C. Sabatti** and E. Dell’Angelica (2005) “Distribution and dynamics of Lamp1-containing endocytic organelles in fibroblasts deficient in biogenesis of lysosome-related organelles complex-3 (BLOC-3),” *Journal of Cell Science* **118**: 5243–5255. ◁

24. **Sabatti, C.** and G. James (2006) “Bayesian sparse hidden components analysis for transcription regulation networks,” *Bioinformatics* **22**: 739–746. ◁

25. Ayers, K., **C. Sabatti** and K. Lange (2006) “Reconstructing ancestral haplotypes with a dictionary model,” *Journal of Computational Biology* **13**: 767–785. ◁

26. Service, S., J. De Young, H. Pretorious, J. Roos, M. Karayiorgou, G. Bedoya, J. Ospina, A. Ruiz Linares, A. Macedo, J. Palha, P. Heutink, Y. Aulchenko, B. Oostra, C. van Duijn, M. Jarvelin, T. Varilo, L. Peltonen, L. Peddle, P. Rahman, G. Piras, M. Monne, S. Murray, L. Galver, **C. Sabatti**, A. Collins and N. Freimer (2006) "Distribution and magnitude of linkage disequilibrium in population isolates and implications for genome-wide association studies," *Nature Genetics* **38**: 556–560. ◁

27. Lee, H. J. Jen, H. Wang, Z. Chen, H. Mamsa, **C. Sabatti**, R. Baloh and S. Nelson (2006), "A genome-wide linkage scan of familial benign recurrent vertigo: linkage to 22q12 with evidence of heterogeneity," *Human Molecular Genetics* **15**: 251–258. ◁

28. Herzberg, I., A. Jasinska, J. García, D. Jawaheer, S. Service, B. Kremeyer, C. Duque, M. Parra, J. Vega, D. Ortiz, L. Carvajal, G. Polanco, G. Restrepo, C. López, C. Palacio, M. Levinson, I. Aldana, C. Mathews, P. Davanzo, J. Molina, E. Fournier, J. Bejarano, M. Ramírez, C. Ortiz, X. Araya, **C. Sabatti**, V. Reus, G. Macaya, G. Bedoya, J. Ospina, N. Freimer and A. Ruiz-Linares (2006) "Convergent linkage evidence from two Latin-American population isolates supports the presence of a susceptibility locus for bipolar disorder in 5q31-34," *Human Molecular Genetics* **15**: 3146–3153. ◁

29. Keen-Kim, D., C. Mathews, V. Reus, T. Lowe, L. Herrera, C. Budman, V. Gross-Tsur, A. Pulver, R. Bruun, G. Erenberg, A. Naarden, **C. Sabatti** and N. Freimer (2006) "Overrepresentation of rare variants in a specific ethnic group may confuse interpretation of association analyses," *Human Molecular Genetics* **15**: 3324–3328. ◁

30. Service, S., J. Molina, J. Deyoung, D. Jawaheer, I. Aldana, T. Vu, J. Bejarano, E. Fournier, M. Ramirez, C. Mathews, P. Davanzo, G. Macaya, L. Sandkuijl, **C. Sabatti**, V. Reus and N. Freimer (2006) "Results of a SNP genome screen in a large Costa Rican pedigree segregating for severe bipolar disorder," *American Journal of Medical Genetics Part B (Neuropsychiatric Genetics)* **141B**: 367–73. ◁

31. Wang, H., C. Lin, S. Service, The international collaborative group on isolated populations, Y. Chen, N. Freimer and **C. Sabatti** (2006) "Linkage disequilibrium and haplotype homozygosity in population samples genotyped at a high marker density," *Human Heredity* **62**: 175–189. ◁

32. Chen, Y., C. Lin and **C. Sabatti** (2006) "Volume measures for linkage disequilibrium," *BMC Genetics* **7**: 54 (electronic edition). ◁

33. Ayers, K., **C. Sabatti** and K. Lange (2007) "A dictionary model for haplotyping, genotype calling, and association testing," *Genetic Epidemiology* **31**: 672–683. ◁

34. Cha, Y., J. Brodsky, G. Ishiyama, **C. Sabatti** and R. Baloh (2007) "The relevance of migraine in patients with Ménière's disease," *Acta Oto-laryngologica* **127**: 1241–1254. ◁

35. Service, S., The international collaborative group on isolated populations, **C. Sabatti** and N. Freimer (2007) "Tag SNPs chosen from HapMap perform well in several population isolates," *Genetic Epidemiology* **31**: 189–194. ◁

36. Cha, Y., J. Brodsky, G. Ishiyama, **C. Sabatti** and R. Baloh (2008) "Clinical features and associated syndromes of mal de debarquement," *Journal of Neurology* **255**: 1038–44. ◁

37. **Sabatti, C.** and K. Lange (2008) "Bayesian Gaussian mixture models for high density genotyping arrays," *Journal of the American Statistical Association* **103**: 89–100. ◁

38. Stefansson, H. [...] **C. Sabatti** [...] K. Stefansson (2008) "Large recurrent microdeletions associated with schizophrenia," *Nature* **455**: 232–236. ◁

39. Vrijenhoek, T., J. Buizer-Voskamp, I. van der Stelt, E. Strengman, Genetic Risk and Outcome in Psychosis (GROUP) Consortium, **C. Sabatti**, A. van Kessel, H. Brunner, R. Ophoff and J. Veltman (2008) "Recurrent CNVs disrupt three candidate genes in schizophrenia patients," *The American Journal of Human Genetics* **83**: 504–510. ◁

40. **Sabatti. C.**, S. Service, A. Hartikainen, A. Pouta, S. Ripatti, J. Brodsky, C. Jones, N. Zaitlen, T. Varilo, M. Kaakinen, U. Sovio, A. Ruokonen, J. Laitinen, E. Jakkula, C. Lachlan, C. Hoggart, P. Elliott, A. Collins, H. Turunen, S. Gabriel, M. McCarthy, M. Daly, M-R. Jarvelin, N. Freimer and L. Peltonen (2009) "Genomewide association analysis of metabolic phenotypes in a birth cohort from a founder population," *Nature Genetics* **41**: 35–46. ◁

41. Aulchenko, Y., S. Ripatti, I. Lindqvist, D. Boomsma, I. Heid, P. Pramstaller, B. Penninx, A. Janssens, J. Wilson, T. Spector, N. Martin, N. Pedersen, K. Kyvik, J. Kaprio, A. Hofman, N. Freimer, M. Jarvelin, U. Gyllensten, H. Campbell, I. Rudan, A. Johansson, F. Marroni, C. Hayward, V. Vitart, I. Jonasson, C. Pattaro, A. Wright, N. Hastie, I. Pichler, A. Hicks, M. Falchi, G. Willemsen, J. Hottenga, E. de Geus, G. Montgomery, J. Whitfield, P. Magnusson, J. Saharinen, M. Perola, K. Silander, A. Isaacs, E. Sijbrands, A. Uitterlinden, J. Witteman, B. Oostra, P. Elliott, A. Ruokonen, **C. Sabatti**, C. Gieger, T. Meitinger, F. Kronenberg, A. Doering, H. Wichmann, J. Smit, M. McCarthy, C. van Duijn and L. Peltonen (2009) "Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts," *Nature Genetics* **41**: 47–55. ◁

42. Wang, H., J. Veldink, H. Blauw, L. van den Berg, R. Ophoff and **C. Sabatti** (2009) "Markov models for inferring copy number variations from genotype data on Illumina platforms," *Human Heredity* **68**: 1–22. ◁

43. Rujescu, D., A. Ingason, S. Cichon, O. Pietiläinen, M. Barnes, T. Toulopoulou, M. Picchioni, E. Vassos, U. Ettinger, E. Bramon, R. Murray, M. Ruggeri, S. Tosato, C. Bonetto, S. Steinberg, E. Sigurdsson, T. Sigmundsson, H. Petursson, A. Gylfason, P. Olason, G. Hardarsson, G. Jonsdottir, O. Gustafsson, R. Fossdal, I. Giegling, H. Möller, A. Hartmann, P. Hoffmann, C. Crombie, G. Fraser, N. Walker, J. Lonnqvist, J. Suvisaari, A. Tuulio-Henriksson, S. Djurovic, I. Melle, O. Andreassen, T. Hansen, T. Werge, L. Kiemeney, B. Franke, J. Veltman, J. Buizer-Voskamp; GROUP Investigators, **C. Sabatti**, R. Ophoff, M. Rietschel, M. Nöthen, K. Stefansson, L. Peltonen, D. St Clair, H. Stefansson and D. Collier (2009) "Disruption of the neurexin 1 gene is associated with schizophrenia," *Human Molecular Genetics* **18**: 988–96. ◁

44. Jasinska, A., S. Service, D. Jawaheer, J. DeYoung, M. Levinson, Z. Zhang, B. Kremeyer, H. Muller, I. Aldana, J. Garcia, G. Restrepo, C. Lopez, C. Palacio, C. Duque, M. Parra, J. Vega, D. Ortiz, G. Bedoya, C. Mathews, P. Davanzo, E. Fournier, J. Bejarano, M. Ramirez, C. Araya Ortiz, X. Araya, J. Molina, **C. Sabatti**, V. Reus, J. Ospina, G. Macaya, A. Ruiz-Linares and N. Freimer (2009) "Narrow and highly significant linkage signal for severe bipolar disorder in the chromosome 5q33 region in latin american pedigrees," (2009) *American Journal of Medical Genetics Part B (Neuropsychiatric Genetics)* **150B**: 998–1006. ◁

45. Hattori, D., Y. Chen, B. Matthews, L. Salwinski, D. Eisenberg, **C. Sabatti**, W. Grueber and L. Zipuski (2009) "Robust discrimination between self and non-self neurites requires thousands of Dscam1 isoforms," *Nature* **461**: 644–648. ◁

46. Ghiani, C., M. Starcevic, I. Rodriguez-Fernandez, R. Nazarian, V. Cheli, L. Chan, J. Malvar, J. de Vellis, **C. Sabatti** and E. Dell'Angelica (2010) "The dysbindin-containing complex (BLOC-1) in brain: developmental regulation, interaction with SNARE proteins, and role in neurite outgrowth," *Molecular Psychiatry* **15**: 204–215. ◁

47. Kang, H., J-H. Sul, S. Service, N. Zaitlen, S. Kong, N. Freimer, **C. Sabatti*** and E. Eskin* (2010) "Variance component model to account for sample structure in genome-wide association studies," *Nature Genetics* **42**: 348–354. * joint corresponding authors. ◁

48. Zhang, Z., K. Lange, R. Ophoff and **C. Sabatti** (2010) "Reconstructing DNA copy number by penalized estimation and imputation," *The Annals of Applied Statistics* **4**: 1749–1773. ◁

49. James, G., **C. Sabatti**, N. Zhou and J. Zhu (2010) "Sparse regulatory networks," *The Annals of Applied Statistics* **4**: 663–686. ◁

50. Teslovich, T. [...] **C. Sabatti** [...] S. Kathiresan (2010) "Biological, clinical and population relevance of 95 loci for blood lipids," *Nature* **466**: 707–713. ◁

51. Pietilainen, O., K. Rehnstrom, E. Jakkula, S. Service, E. Congdon, C. Tilgmann, A. Hartikainen, A. Taanila, U Heikura, T. Paunio, S. Ripatti, M. Jarvelin, M. Isohanni, **C. Sabatti**, A. Palotie, N. Freimer and L. Peltonen (2011) "Phenotype mining in CNV carriers from a population cohort,"*Human Molecular Genetics* **20**: 2686–95. ◁

52. Buizer-Voskamp, J., J. Muntjewerff, Genetic Risk and Outcome in Psychosis (GROUP) Consortium, E. Strengman, **C. Sabatti**, H. Stefansson, J. Vorstman and R. Ophoff (2011) "Genome-wide analysis shows increased frequency of copy number variation deletions in Dutch schizophrenia patients," *Biological Psychiatry* **70**: 655–62. ◁

53. Visnyei, K., H. Onodera, R. Damoiseaux, K. Saigusa, S. Petrosyan, D. De Vries, D. Ferrari, J. Saxe, E. Panosyan, M. Masterman-Smith, J. Mottahedeh, K. Bradley, J. Huang, **C. Sabatti**, I. Nakano and H. Kornblum (2011) "A molecular screening approach to identify and characterize inhibitors of glioblastoma multiforme stem cells," *Molecular Cancer Therapeutics* **10**: 1818–28. ◁

54. Dastani, Z. [...] **C. Sabatti** [...] S. Kathiresan (2012) “Novel loci for adiponectin levels and their influence on type 2 diabetes and metabolic traits: a multi-ethnic meta-analysis of 45,891 individuals,” *PLoS Genetics* **8**: e1002607 (electronic edition). ◁

55. Service, S., K. Verweij, J. Lahti, E. Congdon, J. Ekelund, M. Hintsanen, K. Räikkönen, T. Lehtimäki, M. Kähönen, E. Widen, A. Taanila, J. Veijola, A. Heath, P. Madden, G. Montgomery, **C. Sabatti**, M. Järvelin, A. Palotie, O. Raitakari, J. Viikari, N. Martin, J. Eriksson, L. Keltikangas-Järvinen, N. Wray and N. Freimer (2012) “A genome-wide meta-analysis of association studies of Cloninger's Temperament Scales,” *Translational Psychiatry* **2**: e116 (electronic edition). ◁

56. Zhang, Z., K. Lange and **C. Sabatti** (2012) “Reconstructing DNA copy number by joint segmentation of multiple sequences,” *BMC Bionformatics* **13**: 205 (electronic edition). ◁

57. Scharf, J. [...] **C. Sabatti** [...] D. Pauls (2013) “Genome-wide association study of Tourette's syndrome,” *Molecular Psychiatry* **18**: 721–8. ◁

58. Buizer-Voskamp, J., H. Blauw, M. Boks, K. van Eijk, J. Veldink, E. Hennekam, J. Vorstman, F. Mulder, H. Tiemeier, A. Uitterlinden, L. Kiemeney, L. van den Berg, R. Kahn, **C. Sabatti**, and R. Ophoff (2013) “Increased paternal age and the influence on burden of genomic copy number variation in the general population,” *Hum Genet.* **132**: 443–50. ◁

59. Fears, S., S. Service, T. Teshiba, C. Araya, X. Araya, J. Bejarano, J. Gomez-Franco, B. Kremeyer, Z. Abaryan, I. Aldana, M. Ericson, M. Jalbrzkowski, J. Luykx, L. Navarro, N. Sharif, L. Altshuler, G. Bartzokis, J. Escobar, D. Glahn, J. Ospina-Duque, N. Risch, A. Ruiz-Linares, R. Cantor, C. Lopez-Jaramillo, G. Macaya, J. Molina, V. Reus, **C. Sabatti**, N. Freimer, and C. Bearden (2014) “Multi-system component phenotypes of bipolar disorder for genetic investigations of extended pedigrees” *JAMA Psychiatry* **71**: 375–87. ◁

60. Service, S., T. Teslovich, C. Fuchsberger, V. Ramenksy, P. Yajnik, D. Koboldt; D. Larson, Q. Zhang, L. Lin, R. Welch, L. Ding, M. McLellan, M. O'Laughlin, C. Fronick, L. Fulton; V. Magrini, P. Elliott, M. Jarvelin, M. Kaakinen, M. McCarthy, L. Peltonen, A. Pouta, L. Bonnycastle, F. Collins, N. Narisu, H. Stringham, J. Tuomilehto, S. Ripatti, R. Fulton, **C. Sabatti**, R. Wilson, M. Boehnke, and N. Freimer (2014) “Re-sequencing expands our understanding of the phenotypic impact of variants at GWAS loci,” *PLoS Genetics* **10**: e1004147 (electronic edition). ◁

61. Yu, D. [...] **C. Sabatti** [...] D. Pauls (2015) “Cross-disorder genome-wide analyses suggest a complex genetic relationship between Tourette's syndrome and OCD,” *Am J Psychiatry,* **172**: 82–93. ◁

62. Fears, S., R. Schür, R. Sjouwerman, S. Service, C. Araya, X. Araya, J. Bejarano, E. Knowles, J. Gomez-Makhinson, M. Lopez, I. Aldana, T. Teshiba, Z. Abaryan, N. Al-Sharif, L. Navarro, T. Tishler, L. Altshuler, G. Bartzokis, J. Escobar, D. Glahn, P. Thompson, C. Lopez-Jaramillo, G. Macaya, J. Molina, V. Reus, **C. Sabatti**, R. Cantor, N. Freimer, C. Bearden (2015) “Brain structure-function associations in multi-generational families genetically enriched for bipolar disorder,” *Brain,* **138**: 2087–102. ◁

63. Bogdan, M., E. van den Berg, **C. Sabatti**, W. Su, E. Candes (2015) “SLOPE – Adaptive variable selection via convex optimization,” *Annals of Applied Statistics*, **9**: 1103–40 ◁

64. Banda, Y., M. Kvale, T. Hoffmann, S. Hesselson, D. Ranatunga, H. Tang, **C. Sabatti**, L. Croen, B. Dispensa, M. Henderson, C. Iribarren, E. Jorgenson, L. Kushi, D. Ludwig, D. Olberg, C. Quesenberry, S. Rowell, M. Sadler, L. Sakoda, S. Sciortino, L. Shen, D. Smethurst, C. Somkin, S. Van Den Eeden, L. Walter, R. Whitmer, P. Kwok, C. Schaefer and N. Risch (2015) "Characterizing race/ethnicity and genetic ancestry for 100,000 subjects in the Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort," *Genetics*, **200**: 1285–95. ◁

65. Peterson, C., M. Bogomolov, Y. Benjamini and **C. Sabatti** (2016) "Many phenotypes without many false discoveries: error controlling strategies for multi-traits association studies,"*Genetic Epidemiology*, **40**: 45–56. ◁

66. Stell, L. and **C. Sabatti** (2016) "Genetic variant selection: learning across traits and sites," *Genetics* **202**: 439–55. ◁

67. Pagani, L., P. St. Clair, T. Teshiba, S. Service, S. Fears, C. Araya, X. Araya, J. Bejarano, M. Ramirez, G. Castrillon, J. Gomez-Makhinson, M. Lopez, G. Montoya, P. Montoya, I. Aldana, L. Navarro, D. Freimer, B. Safaie, L. Keung, K. Greenspan, K. Chou, J. Escobar, J. Ospina-Duque, B. Kremeyer, A. Ruiz, R. Cantor, C. Lopez-Jaramillo, G. Macaya, J. Molina, V. Reus, **C. Sabatti**, C. Bearden, J. Takahashi and N. Freimer (2016) "Genetic contributions to circadian rhythm and sleep phenotypes in pedigrees segregating for severe bipolar disorder," *PNAS* **113**: E754–61. ◁

68. Peterson, C., M. Bogomolov, Y. Benjamini and **C. Sabatti** (2016) "TreeQTL: hierarchical error control for eQTL findings," *Bioinformatics*, **32**: 2556–8. ◁

69. Peterson, C., S. Service, A. Jasinska, F. Gao, I. Zelaya, T. Teshiba, C. Bearden, V. Resus, G. Macaya, C. Lopez, M. Bogomolov, Y. Benjamini, E. Eskin, G. Coppola, N. Freimer, and **C. Sabatti** (2016) "Genetic regulation of LCL gene expression in families segregating bipolar disorder," *PLOS Genetics*, **12**:e1006046 ◁

70. Brzyski, D., C. Peterson, P. Sobczyk, E. Candes, M. Bogdan, **C. Sabatti** (2017) "Controlling the rate of GWAS false discoveries," *Genetics* **205**: 61–75.◁

71. GTEx Consortium, Lead analysts, Laboratory, Data Analysis and Coordinating Center, Biospecimen collection, Pathology, **eQTL manuscript working group**, A. Battle, C. Brown, B. Engelhart, S. Montgomery (2017) "Genetic effects on gene expression across human tissues human tissue," *Nature*, 550: 204–213 ◁

72. Jasinska, A., I. Zelaya, S. Service, C. Peterson, R. Cantor, O. Choi, J. DeYoung, E. Eskin, L. Fairbanks, S. Fears, A. Furterer, Y. Huang, V. Ramensky, C. Schmitt, H. Svardal, M. Jorgensen, J. Kaplan, D. Villar, B. Aken, P. Flicek, R. Nag, E. Wong, J. Blangero, T. Dyer, M. Bogomolov, Y. Benjamini, G. Weinstock, K. Dewar, **C. Sabatti**, R. Wilson, J. Jentsch, W. Warren, G. Coppola, R. Woods, N. Freimer (2017) "Genetic variation and gene expression across multiple tissues and developmental stages in a nonhuman primate," *Nature Genetics*, **49**: 1714–1721. ◁

73. Sanders, S., [...] **C. Sabatti**, [...] N. Freimer (2017) “Whole genome sequencing in psychiatric disorders: the WGSPD consortium,” *Nat Neurosci*, **20**: 1661–1668. ◁

74. Prunicki, M., L. Stell, D. Dinakarpandian, M. de Planell-Saguer, R. Lucas, S. Hammond, J. Balmes, X. Zhou, T. Paglino, **C. Sabatti**, R. Miller, K. Nadeau (2018) “Exposure to NO$_2$, CO, and PM$_{2.5}$ is linked to regional DNA methylation differences in asthma,” *Clin Epigenetics*, 10:2. ◁

75. Mooney, J., C. Huber, S. Service, J. Sul, C. Marsden, Z. Zhang, **C. Sabatti**, A. Ruiz-Linares, G. Bedoya; Costa Rica/Colombia Consortium for Genetic Investigation of Bipolar Endophenotypes, N. Freimer, K. Lohmueller (2018) “Understanding the Hidden Complexity of Latin American Population Isolates”, *Am. J. Hum. Genet.* **103**: 707–726. ◁

76. Neal, J., X. Li, J. Zhu, V. Giangarra, C. Grzeskowiak, J. Ju, I. Liu, S. Chiou, A. Salahudeen, A. Smith, B. Deutsch, L. Liao, A. Zemek, F. Zhao, K. Karlsson, L. Schultz, T. Metzner, L. Nadauld, Y. Tseng, S. Alkhairy, C. Oh, P. Keskula, D. Mendoza-Villanueva, F. De La Vega, P. Kunz, J. Liao, J. Leppert, J. Sunwoo, **C. Sabatti**, J. Boehm, W. Hahn, G. Zheng, M. Davis, C. Kuo (2018) “Organoid Modeling of the Tumor Immune Microenvironment,” *Cell*, **175**: 1972–1988. ◁

77. Katsevich, G. and **C. Sabatti** (2019) “Multilayer Knockoff Filter: Controlled variable selection at multiple resolutions,” *Annals of Applied Statistics*, **13**: 1–33. ◁

78. Sesia, M., **C. Sabatti** and E. J. Candès (2019) “Gene hunting with hidden Markov model knockoffs,” *Biometrika*, **106**: 1–18. ◁

79. C. Hoerner, R. Massoudi, T. Metzner, L. Stell, J. O’Rourke, C. Kong, J. Liliental, J. Brooks, **C. Sabatti**, J. Leppert, A. Fan (2019) “Multiregion Quantification of Extracellular Signal-regulated Kinase Activity in Renal Cell Carcinoma,” *European Urology Oncology* S2588-9311(18)30171-8 ◁

80. Zhu, J. Q. Zhao, E. Katsevich, **C. Sabatti** (2019) “Exploratory Gene Ontology Analysis with Interactive Visualization,” *Scientific Reports* **24**: 7793. ◁

81. Wojcik, G. [...] **C. Sabatti** [...] C. Carlson (2019) “The PAGE Study: How Genetic Diversity Improves Our Understanding of the Architecture of Complex Traits,” *Nature* **570**: 514–518. ◁

82. Locke, A., [...], **C. Sabatti**, [...] N. Freimer (2019) “Exome sequencing identifies high impact trait-associated alleles enriched in Finns,” *Nature* **572**: 323–328. ◁

83. Panigrahi, S., J. Zhu, **C. Sabatti** (2019) “Selection-adjusted inference: an application to confidence intervals for cis-eQTL effect sizes,” *Biostatistics*, Epub ahead of print. ◁

84. Balliu, B., M. Durrant, O. de Goede, N. Abell, X. Li, B. Liu, M. Gloudemans, N. Cook, K. Smith, M. Pala, F. Cucca, D. Schlessinger, S. Jaiswal, **C. Sabatti**, L. Lind, E. Ingelsson, S. Montgomery (2019) “Genetic dysregulation of gene expression and splicing during a ten-year period of human aging,” *Genome Biology* **20**:230. ◁

85. Vreeker, A., S. Fears, S. Service, L. Pagani, J. Takahashi, C. Araya, X. Araya, J. Bejarano, M. Lopez, G. Montoya, C. Montoya, T. Teshiba, J. Escobar, R. Cantor, C. López-Jaramillo, G. Macaya, J. Molina, V. Reus, **C. Sabatti**, R. Ophoff, N. Freimer, C. Bearden (2019) "Genetic analysis of activity, brain and behavioral associations in extended families with heavy genetic loading for bipolar disorder," *Psychol Med.* ◁

86. Sesia, M., E. Katsevich, S. Bates, E. Candès and **C. Sabatti** (2020) "Multi-resolution localization of causal variants across the genome" *Nature Communications* **11**: 1–10. ◁

87. Sul, J. [...] **C. Sabatti**, N. Freimer (2020) "Contribution of common and rare variants to bipolar disorder susceptibility in extended pedigrees from population isolates," *Transl Psychiatry* **10**:74. ◁

88. Romano, Y., R. Barber, **C. Sabatti**, E, Candès (2020) "With Malice Towards None: Assessing Uncertainty via Equalized Coverage," *Harvard Data Science Review* **2.2** ◁

89. Service, S., M. Umanes, C. Upegui, M. Ramírez, L. Arango, A. Díaz-Zuluaga, Juanita Melo Espejo, M. López, J. Palacio, S. Ruiz, J. Valencia, T. Teshiba, A. Espinoza, L. Loohuis, B. Brodey, **C. Sabatti**, T. Moore, A. Port, J. Escobar, V. Reus, C. Jaramillo, R. Gur, C. Bearden, N. Freimer (2020) "Distinct and shared contributions of diagnosis and symptom domains to cognitive performance in severe mental illness" *Lancet Psychiatry* **7**: 411–419. ◁

90. S. Fears, S. Service, B. Kremeyer, C. Araya, X. Araya, J. Bejarano, M. Ramirez, G. Castrillón, J. Gomez-Franco, M. Lopez, G. Montoya, P. Montoya, I. Aldana, T. Teshiba, N. Al-Sharif, M. Jalbrzikowski, T. Tishler, J. Escobar, A. Ruiz-Linares, C. Lopez-Jaramillo, G. Macaya, J. Molina, V. Reus, R. Cantor, **C. Sabatti**, N. Freimer, C. Bearden (2020) "Genome-wide mapping of brain phenotypes in extended pedigrees with strong genetic loading for bipolar disorder," *Molecular Psychiatry* ◁

91. Bogomolov, M., C. Peterson, Y. Benjamini, and **C. Sabatti** (2021) "Testing hypotheses on a tree: new error rates and controlling strategies," *Biometrika*, **108**: 575–590, ◁

92. Bates, S., M. Sesia, **C. Sabatti**, and E. Candès (2020) "Causal inference in genetic trio studies" *PNAS*, **117** 24117–24126. ◁

93. Salahudeen, A., S. Choi, A. Rustagi, J. Zhu, V. van Unen, S. de la O, R. Flynn, M. Margalef-Català, A. Santos, J. Ju, A. Batish, T. Usui, G. Zheng, C. Edwards, L. Wagar, V. Luca, B. Anchang, M. Nagendran, K. Nguyen, D. Hart, J. Terry, P. Belgrader, S. Ziraldo, T. Mikkelsen, P. Harbury, J. Glenn, K. Garcia, M. Davis, R. Baric, **C. Sabatti**, M. Amieva, C. Blish, T. Desai, C. Kuo (2020) "Progenitor identification and SARS-CoV-2 infection in human distal lung organoids." *Nature*, **588**: 670–675. ◁

94. Kalinowski, A., J. Liliental, L. Anker, O. Linkovski, C. Culbertson, J. Hall, R. Pattni, **C. Sabatti**, D. Noordsy, J. Hallmayer, E. Mellins, J. Ballon, R. O'Hara, D. Levinson, A. Urban (2021) "Increased activation product of complement 4 protein in plasma of individuals with schizophrenia," *Transl Psychiatry* **11**: 486. ◁

95. Markin, C., D Mokhtari, F. Sunden, M. Appel, E. Akiva, S. Longwell, **C. Sabatti**, D. Herschlag, P. Fordyce (2021) “Revealing enzyme functional architecture via high-throughput microfluidic enzyme kinetics,” *Science*, **373**: eabf8761 ◁

96. Sesia, M., S. Bates, E. Candès, J. Marchini, and **C. Sabatti** (2021) “False discovery rate control in genome-wide association studies with population structure” *PNAS* **118**: e2105841118 ◁

97. Li, S., M. Sesia, Y. Romano, E. Candès, **C. Sabatti** (2022) “Searching for consistent associations with a multi-environment knockoff filter,” *Biometrika* **1-09**: 611–629. ◁

98. Wang, J., Gui, L., W. Su, **C. Sabatti**, A. Owen (2022) “Detecting Multiple Replicating Signals using Adaptive Filtering Procedures,” *The Annals of Statistics*, **50**: 1890–1909. ◁

99. Li, S., Z. Ren, **C. Sabatti**, M. Sesia (2022) “Transfer learning in genome-wide association studies with knockoffs,” *Sankhya B*. ◁

100. He, Z., L. Liu, M. Belloy, Y. Le Guen, A. Sossin, X. Liu, X. Qi, S. Ma, T. Wyss-Coray, H. Tang, **C. Sabatti**, E. Candès, M. Greicius, I. Ionita-Laza (2022) “Summary statistics knockoff inference empowers identification of putative causal variants in genome-wide association studies,” *Nature Communications*, **13**: 7209. ◁

101. Katsevich, E., **C. Sabatti**, M. Bogomolov (2023) “Filtering the rejection set while preserving false discovery rate control,” *JASA*, **118**: 165–176. ◁

102. J. Song, M. Castano Ramírez, J. Okano, S. Service, J. de la Hoz, A. ìaz-Zuluaga, C. Vargas Upegui, C. Gallago, A. Arias, A. Sánchez, T. Teshiba, **C. Sabatti**, R. Gur, C. Bearden, J. Escobar, V. Reus, C. Jaramillo, N. Freimer, L. Olde Loohuis, and S. Blower. (2024) “Geospatial investigations in colombia reveal variations in the distribution of mood and psychotic disorders,” *Communications Medicine* bf 4: article number 26. ◁

103. Chu, B., J. Gu, Chen, Z., T. Morrison, E. Candès, Z. He and **C. Sabatti** (2024) “Second-order group knockoffs with applications to GWAS,” *Bioinformatics* ◁

104. Service, S., J. De La Hoz, A. Diaz-Zuluaga, A. Arias, A. Pimplaskar, C. Luu, L. Mena, J. Valencia, M. Castano Ramírez, C. Bearden, **C. Sabatti**, V. Reus, C. López-Jaramillo, N. Freimer, L. Olde Loohuis (2025) “Predicting diagnostic conversion from major depressive disorder to bipolar disorder: an EHR based study from Colombia,” *Bipolar Disorder* **27**:47–56.◁

105. Gablenz, P. and **C. Sabatti** (2025) “Catch me if you can: Signal localization with knockoff e-values,” *JRSS B* **87**:56–73. ◁

106. **Sabatti, C.** and Q. Zhao (2025) “Near-peer mentoring in data science: a plot for mutual growth ” *The American Statistician*, **79**:529–537. ◁

Research Papers - submitted to journals

107. Chen, Z., Z. He, B. Chu, J. Gu, T. Morrison, **C. Sabatti** and E Candès (2024) "Controlled Variable Selection from Summary Statistics Only? A Solution via GhostKnockoffs and Penalized Regression," arXiv:2402.12724, revised for *The Annals of Applied Statistics* ◁

108. He, Z., B. Chu, J. Yang, J. Gu, Chen, Z., L. Liu, T. Morrison, M. Belloy, X. Qi, N. Hejazi, M. Mathur, Y. LeGuenn, H. Tang, T. Hastie, I. Ionita-Iaza, E. Candès and **C. Sabatti** (2024) "Beyond guilty by association at scale: searching for causal variants on the basis of genome-wide summary statistics," bioRxiv 2024.02.28.582621, in revision for *Genome Biology* ◁

109. Gablenz, P., M. Sesia, T. Sun and **C. Sabatti** (2024) "Searching for local associations while controlling the false discovery rate," arXiv:2412.02182, revised for *Journal of the American Statistical Association* ◁

110. Boltz, T.*, B. Chu*, C. Liao*, J. Sealock, R. Ye, L. Majara, J. Fu, S. Service, L. Zhan, S. Medland, S. Chapman, S. Rubinacci, M. DeFelice, J. Grimsby, T. Abebe, M. Alemayehu, F. Ashaba, E. G Atkinson, T. Bigdeli, A. Bradway, H. Brand, L. Chibnik, A. Fekadu, M. Gatzen, B. Gelaye, S. Gichuru, M. Gildea, T. Hill, H. Huang, K. Hubbard, W. Injera, R. James, M. Joloba, C. Kachulis, P. Kalmbach, R. Kamulegeya, G. Kigen, S. Kim, N. Koen, E. Kwobah, J. Kyebuzibwa, S. Lee, N. Lennon, P. Lind, E. Lopera-Maya, J. Makale, S. Mangul, J. McMahon, P. Mowlem, H. Musinguzi, R. Mwema, N. Nakasujja, C. Newman, L. Nkambule, C. O'Neil, A. Olivares, C. Olsen, L. Ongeri, S. Parsa, A. Pretorius, R. Ramesar, F. Reagan, **C. Sabatti**, J. Schneider, W. Shiferaw, A. Stevenson, E. Stricker, R. Stroud, J. Tang, D. Whiteman, M. Yohannes, M. Yu, K. Yuan, NeuroGAP-Psychosis, D. Akena, L. Atwoli, S. Kariuki, K. Koenen, C. Newton, D. Stein, S. Teferra, Z. Zingela, C. Pato, M. Pato, C.Lopez-Jaramillo, N. Freimer, R. Ophoff, L. Loohuis, M. Talkowski, B. Neale, D. Howrigan, A. Martin (2024) "A blended genome and exome sequencing method captures genetic variation in an unbiased, high-quality, and cost-effective manner," bioRxiv 2024.09.06.611689 ◁

111. Liu, X. J. Gu, Z. Chen, B. Chu, L. Liu, T. Morrison, R. Butler, J. Edelson, J. Li, F. Longo, H. Tang, I. Ionita-laza, **C. Sabatti***, E. Candès*, Z. He* (2025) "Uncovering Heterogeneous Effects via Localized Feature Selection," bioRxiv 2025.06.03.657761, in revision for *PNAS* ◁

112. Chu, B., Z. He, and **C. Sabatti** (2025) "It's a wrap: deriving distinct discoveries with FDR control after a GWAS pipeline," bioRxiv 2025.06.05.658138 ◁

113. Zhao, Q., S. Service, C. Bearden, C. Lopez-Jaramillo, N. Freimer and **C. Sabatti** (2025) "Mapping beyond diseases: Controlled variable selection for secondary phenotypes using tilted knockoffs," arXiv:2508.18548 ◁

Research Papers - Peer reviewed (Conference Proceedings)

114. Liu, J. and **C. Sabatti** (1999) "Simulated sintering: Markov chain Monte Carlo with spaces of varying dimensions," *Bayesian Statistics 6*, J.M. Bernardo, J.O. Berger, A. P. Dawid, and A.F.M. Smith (eds). New York: Oxford University Press.

115. **Sabatti, C.**, L. Rohlin and J. Liao (2003) "Dictionary model for the analysis of E. Coli promoter regions," *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* **4**: 3711–3714. ◁

116. **Sabatti, C.** and L. Rohlin (2004) "A Bayesian approach to expression network component analysis," *Proceedings of the 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* **2**: 2933–2936. ◁

Chapters

117. **Sabatti, C.** (2006) "False discovery rate and multiple comparisons procedures" *DNA microarrays and related genomic techniques*, Allison, Page, Beasley, and Edwards ed., Chapman & Hall/CRC, pg. 289–304.

118. **Sabatti, C.** (2007) "Avoiding false discoveries in association studies," *Linkage disequilibrium and association mapping: analysis and applications*, A. Collins ed., Humana Press, pg.195–211. ◁

119. **Sabatti, C.** (2013) "Multivariate linear models for GWAS," *Advances in Statistical Bioinformatics*, K. Do, S. Qin, M. Vannucci, ed., Cambridge University Press.

Letters to Editor (peer reviewed)

120. **Sabatti, C.** and N. Risch (2003) "Response to the letter 'Gametic and zygotic associations' by Rong-Cai Yang," *Genetics* **165**: 451–452. ◁

Reviews (peer reviewed)

121. **Sabatti, C.** (2002) "Statistical Issues in DNA Microarrays," *Current Genomics* **3**: 7–12. ◁

122. Liao, J. and **C. Sabatti** (2002) "Gene expression arrays," *American Society for Microbiology News* **68**: 432–437.

123. **Sabatti, C.** (2004) "Interplay in the computational analysis of genomic regulatory sequences and microarray expression data," *Current Genomics* **5**: 439–442.◁

Editorials (peer reviewed)

124. Freimer, N. and **C. Sabatti** (2003) "The human phenome project," *Nature Genetics* **34**: 15–21. ◁

125. Freimer, N. and **C. Sabatti** (2004) "The use of pedigree, sib-pair and association studies of common diseases for genetic mapping and epidemiology," *Nature Genetics* **36**: 1045–1051. (Invited contribution.) ◁

126. Freimer, N. and **C. Sabatti** (2005) "Guidelines for association mapping studies in Human Molecular Genetics," *Human Molecular Genetics* **14**: 2481–2483. ◁

127. **Sabatti, C.** (2005) "Comment on the 'Semilinear high-dimensional model for normalization of microarray data: a theoretical analysis and partial consistency' by Fang *et al.*," *Journal of the American Statistical Association* **100**: 796–799. (Invited contribution.)

128. **Sabatti, C.** (2006) "Comment on the 'Likelihood-Based Inference on haplotype effects in genetic association studies' by Lin and Zeng," *Journal of the American Statistical Association* **101**: 104–106. (Invited contribution.)

129. Freimer, N. and **C. Sabatti** (2007) "Human genetics: variants in common diseases." *Nature* **445**: 828–30. (Invited contribution.) ◁

130. Candès, E., J. Duchi, **C. Sabatti** (2019) "Statistics and the oncoming AI revolution," *Harvard Data Science Review*, Epub ◁

131. **C. Sabatti** and J. Chambers (2022) "Data Science in a Time of Crisis: Lessons from the Pandemic" *Statistical Science*, **37**: 160–61. ◁

132. Yeh, C., D. Wall, K. Matthys, **C. Sabatti** and J. Palacios (2025) "Curriculum Design in an Evolving Field: Perspectives on Biomedical Data Science from Stanford," Annual Review of Biomedical Data Science **8**. ◁

Book reviews

133. **Sabatti, C.** (2004) "Review of 'The analysis of gene expression data: methods and software.' Giovanni Parmigiani, Elizabeth S. Garrett, Rafael A. Irizarry and Scott L. Zeger (eds), Springer, New York, 2003." *Statistics in Medicine* **23**: 3245–3246.

134. **Sabatti, C.** (2007) "Review of 'The statistics of gene mapping.' David Siegmund and Benjamin Yakir, Springer-Verlag, New York, 2007." *Journal of Statistical Software* **21** (electronic edition).

135. **Sabatti, C.** (2009) "Review of 'Applied Statistical Genetics with R for Population-Based Association Studies.' Andrea S. Foulkes, Springer-Verlag, New York, 2009." *Journal of Statistical Software* **31** (electronic edition).

Journal Publications that list C. Sabatti as collaborator

The investigation of the genetic bases of medically relevant traits often requires the creation of consortia of research teams, who pool together samples they have separately collected and analyzed during the years. In the resulting publications, in the effort to limit the number of authors, members of the research teams whose work enabled the final analyses are sometimes recognized as collaborators.

136. Grados, M. *et al.* (2008) "Latent class analysis of gilles de la tourette syndrome using comorbidities: clinical and genetic implications," *Biological Psychiatry* **64**: 219–25.

137. Scharf, J. *et al.* (2008) "Lack of association between SLITRK1var321 and Tourette syndrome in a large family-based sample," *Neurology* **70**: 1495–6.

138. Knight, S. *et al.* (2010) "Linkage analysis of Tourette syndrome in a large Utah pedigree," *American Journal of Medical Genetics Part B (Neuropsychiatric Genetics)* **153B**: 656–62.

139. Crane, J. *et al.* (2011) "Family-based genetic association study of DLGAP3 in Tourette Syndrome," *American Journal of Medical Genetics Part B (Neuropsychiatric Genetics)* **156B**: 108–14.

140. Chambers, J. *et al.* (2011) "Genome-wide association study identifies loci influencing concentrations of liver enzymes in plasma," *Nature Genetics* **43**: 1131–8.

Unpublished reports

141. **Sabatti, C.** (1998) "Group transformations and dimensionality reduction in transition rules for MCMC," Ph.D. Thesis, Statistics Department, Stanford University.

142. **Sabatti, C.**, K. Visnyei and H. Kornblum (2008) "Statistical challenges in high-throughput Screens," *UCLA Statistics Preprint* # 532.