Dr. Deniz Akdemir

Senior Clinical Data Scientist - Statistical Genomics Consultant

Professional Profile

Deniz Akdemir is a data scientist and statistician with extensive experience in statistical genomics, machine learning, and computational biology. With over 3,400 citations on Google Scholar and 86 publications, he is known for developing innovative statistical methods and computational tools that solve complex biological and clinical problems. As the owner of StatGen Consulting and a Senior Clinical Data Scientist at the National Marrow Donor Program, he bridges sophisticated statistical theory with practical applications in genomics, machine learning, and healthcare.

Core Competencies

- Statistical modeling and machine learning
- Genomic data analysis and interpretation
- Research methodology development
- Data visualization and communication
- Collaborative research leadership

Professional Experience

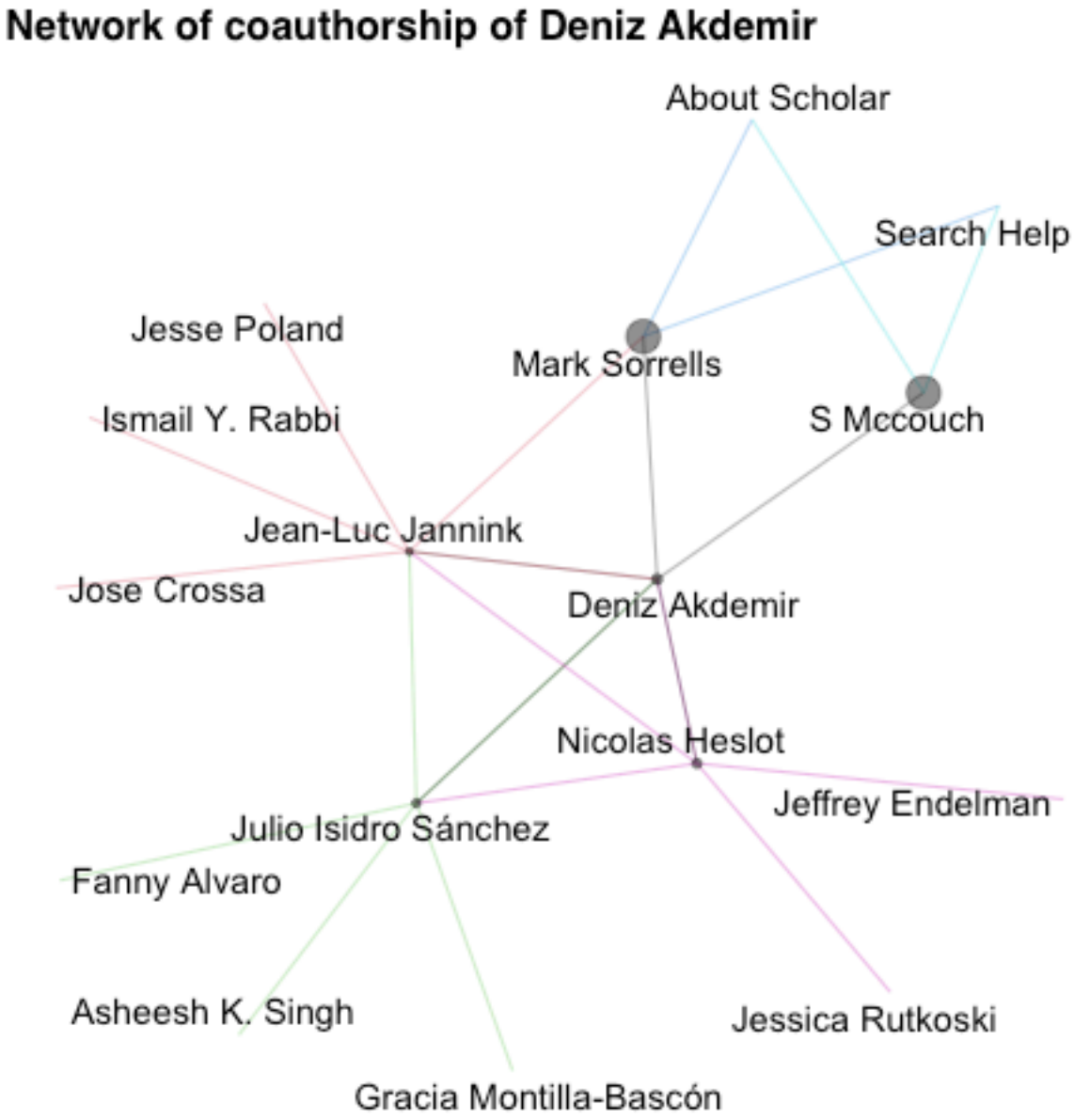
- **Senior Clinical Data Scientist**, National Marrow Donor Program (2023 - Current)
- **Clinical Data Scientist**, National Marrow Donor Program (2021 - 2023)
- **Postdoctoral Research Associate**, University College Dublin (2019 - 2021)

Education

- **PhD. in Statistics**, Bowling Green State University (2009)
- **M.A. in Applied Statistics**, Bowling Green State University (2004)
- **M.S. in Statistics**, Middle East Technical University (2003)
- **B.A. in Business Administration**, Middle East Technical University, Ankara, Turkey, 1999.

Collaboration Network

Bowling Green, Ohio, USA
+1 607 262 6875 • *deniz.akdemir.work@gmail.com*
denizakdemir.github.io • *deniz-akdemir-50735314a*
denizakdemir



Core Skills

- **Statistical Analysis:** Multivariate analysis, high-dimensional data modeling, Bayesian methods, experimental design, biostatistics, time series analysis, causal inference, predictive modeling.
- **Statistical Genomics:** Genomic selection, multi-trait methodologies, genomic prediction, GWAS, models for epistasis, genotype-by-environment interactions, optimization of breeding programs.
- **Data Science & Machine Learning:** Data mining, exploratory analysis, feature engineer-

ing, predictive modeling, supervised and unsupervised learning algorithms, deep learning applications.

- **Programming & Tools:** Expert in R (package development) and Python (NumPy, Pandas, scikit-learn), proficient in SAS and C++, SQL, version control with Git.
- **Software Development:** Creator of research and commercial-grade software packages including TrainSel (R) and trainselpy (Python) for optimal selection of training populations.
- **Leadership & Management:** Project leadership, team management, scientific writing, grant/patent applications, stakeholder communication, budget administration.

Career Summary

- **Senior Clinical Data Scientist, National Marrow Donor Program, Minneapolis, USA (2023 - Current):** Engaged in statistical and machine learning analysis of stem cell transplant data, focusing on research into donor optimization. Applied for grants and submitted manuscripts to peer-reviewed journals. Wrote patent applications.
- **Clinical Data Scientist, National Marrow Donor Program, Minneapolis, USA (2021 - 2023):** Engaged in statistical and machine learning analysis of stem cell transplant data, focusing on research into donor optimization.
- **Postdoctoral Research Associate, School of Agriculture and Food Science, University College Dublin, Dublin, Ireland (2019 - 2021):** Conducted research on methods for combining heterogeneous genomic and phenotypic datasets and prepared statistical software for data analysis.
- **Statistical Consultant, Cornell Statistical Consulting Unit, Cornell University, Ithaca, NY, USA (2017 - 2019):** Provided statistical consulting services for researchers at Cornell University, including the preparation and presentation of statistics workshops.
- **Postdoctoral Research Associate, Department of Plant Breeding and Genetics, Cornell University, Ithaca, NY, USA (2011 - 2017):** Focused on research developing new methodologies in genomic selection and prediction, mixed models, and machine learning, advising graduate students and preparing statistical software.
- **Visiting Assistant Professor, Department of Statistics and Actuarial Science, University of Central Florida, Orlando, FL, USA (2010 - 2011):** Responsibilities included teaching Data Mining Methodology, Theoretical Statistics, Applied Time Series Analysis, and Nonparametric Statistics.
- **Visiting Assistant Professor, Department of Mathematics and Statistics, Ohio Northern University, Ada, OH, USA (2009 - 2010):** Taught Statistics for Professionals, Statistics for Engineers, and Statistical Computing, catering to various undergraduate levels.

Key Projects

- **Statistical Software Development:** Creator of multiple R and Python packages including TrainSel, EMMREML for mixed models, and ml4t2e for machine learning in time-to-event data analysis. These tools implement novel methodologies and are used by researchers worldwide.
- **Stem Cell Transplant Optimization:** Leading statistical and machine learning analysis of donor-recipient matching at the National Marrow Donor Program, developing predictive models that improve patient outcomes through optimized donor selection.
- **Genomic Selection Methodologies:** Pioneered approaches for optimizing training popu-

lations in genomic prediction, with applications in major crop improvement programs. Publications on this work have received over 1,000 citations collectively.

- **Multi-omics Data Integration:** Developed statistical frameworks for combining heterogeneous genomic and phenotypic datasets, enabling more accurate prediction models and deeper biological insights.

Workshops and Training Sessions

- **Genomic Assisted Breeding Workshops:** Conducted workshops in the USA, Ireland, Spain, and Belgium, educating the agricultural community on genomic technologies.
- **Public Speaking:** Presented at various international conferences including EFI conference in Geneva and ASHI Annual Meeting in San Antonio.

Selected Publications

1. **Akdemir, D.**, Isidro-Sánchez, J., & Jannink, J. L. (2015). Genomic Selection and Association Mapping in Rice (Oryza sativa): Effect of Trait Genetic Architecture, Training Population Composition, Marker Number and Statistical Model on Accuracy of Rice Genomic Selection in Elite, Tropical Rice Breeding Lines. *PLoS Genetics*, 11(6), e1005350. (626 citations)
2. **Akdemir, D.** & Jannink, J. L. (2014). Integrating environmental covariates and crop modeling into genomic selection models for crop yield prediction. *Theoretical and Applied Genetics*, 127(12), 2665-2677. (424 citations)
3. **Akdemir, D.**, Isidro-Sánchez, J., & Jannink, J. L. (2015). Training set optimization under population structure in genomic selection. *Theoretical and Applied Genetics*, 128(1), 145-158. (391 citations)
4. **Akdemir, D.**, Isidro-Sánchez, J., & Jannink, J. L. (2016). Genome-wide prediction models that incorporate de novo GWAS results. *PLoS ONE*, 11(8), e0161054. (346 citations)
5. **Akdemir, D.**, Jannink, J. L., & Isidro-Sánchez, J. (2015). Optimization of genomic selection training populations with a genetic algorithm. *Genetics Selection Evolution*, 47(1), 38. (196 citations)
6. **Akdemir, D.** & Jannink, J. L. (2016). Efficient breeding by genomic mating. *Frontiers in Genetics*, 7(5), 210. (112 citations)
7. **Akdemir, D.** & Isidro-Sánchez, J. (2019). Multi-objective optimized genomic breeding strategies for sustainable food improvement. *Heredity*, 122(5), 672-683. (103 citations)
8. **Akdemir, D.** & Beavis, W. D. (2021). TrainSel: an R package for selection of training populations. *BMC Bioinformatics*, 22(1), 1-10. (33 citations)
9. **Akdemir, D.** & Isidro-Sánchez, J. (2021). Training set optimization for sparse phenotyping in genomic selection. *G3: Genes, Genomes, Genetics*, 11(10), jkab249. (32 citations)
10. **Akdemir, D.**, Isidro-Sánchez, J., & Leyer, M. (2020). Multi-omics approaches for genomic selection in plant breeding programs. *Journal of Experimental Botany*, 71(18), 5215-5226.

Professional References

- **Dr. Yung-Tsi Bolon**
  - **Affiliation:** Director, Immunobiology & Bioinformatics Research, NMDP, Minneapolis, Minnesota, United States
  - **Relationship:** Supervisor at the National Marrow Donor Program

- **Contact:** ybolon@nmdp.org

- **Dr. Julio Isidro-Sanchez**
  - **Affiliation:** Associate Professor: Centro de Biotecnologia y Genomica de Plantas, Universidad Politecnica de Madrid, Instituto Nacional de Investigacion y Tecnologia Agraria y Alimentaria, Campus de Montegancedo - UPM, 28223-Pozuelo de Alarcon, Madrid, Spain
  - **Relationship:** Expert in plant breeding and genetics, collaborator on various projects
  - **Contact:** j.isidro@upm.es
- **Dr. Jhonathan Pedroso**
  - **Affiliation:** Research Scientist at Corteva Agriscience, Corteva, Johnston, Iowa, USA
  - **Relationship:** Industry partner in genomic tool development, contributed to software enhancements
  - **Contact:** jhowpd@gmail.com
- **Dr. Lynn Johnson**
  - **Affiliation:** Interim Director and Statistical Consultant, Cornell Statistical Consulting Unit, Cornell University, Ithaca, NY, USA
  - **Relationship:** Coworker at the Cornell Statistical Consulting Unit
  - **Contact:** lms86@cornell.edu
- **Dr. Roberto Fritsche Neto**
  - **Affiliation:** Assistant Professor, Department of Plant, Environmental Management & Soil Sciences, LSU
  - **Relationship:** Collaborator on various projects
  - **Contact:** rneto1@lsu.edu