Daniel Dvorkin
1528 Rosemary St., Denver, CO 80220
(303)548-4568 · daniel.dvorkin@gmail.com

SUMMARY

Experienced biomedical researcher with knowledge in genomics, proteomics, clinomics, statistics, machine learning, decision support, data mining, software engineering, and database architecture and administration. I have authored and co-authored multiple academic publications and software packages and worked as a team lead and team member on internal scientific and computational projects in both academia and industry.

My extensive clinical background has given me a deep understanding of the need to translate biological discoveries into bedside treatments. Ten years of military service taught me how to both serve and lead in critical situations and fast-paced environments. I hope to find a position that will allow me to use my broad range of skills to increase scientific knowledge and improve the efficiency and quality of medical care.

RESEARCH STATEMENT

My primary research interest is in integrative modeling for both clinical and "omics" data, with a focus on human responses and adapatations to acute hypoxia. Combining classical biostatistics and statistically rigorous machine learning methods, I build unified models for a wide variety of data that allow identification of the key players in health and disease processes. Methodologically, I have a particular interest in probabilistic graphical approaches which can simultaneously model biological processes at the genomic, cellular, and phenotypic levels, and in causal inference on such models. As well as mountain sickness, areas of application have included COPD, asthma, and aspiration in humans, and developmental biology in model organisms. I strive to present clear interpretations of analytical results for the benefits of clinicians and biological researchers.

EDUCATION

Ph.D., Computational Bioscience, University of Colorado   2013
Dissertation concerned integration of multiple sources of genome-scale data; see "Publications," "Conference Presentations," and "Research" below. Coursework included general bioinformatics, core biological sciences, statistics and algorithms in genetics and genomics, and mathematical statistics.

M.S., Biostatistics, University of Minnesota   2007
Master's project concerned prediction of gene coexpression using the Gene Ontology; see "Research" below. Coursework included theory of statistics, linear models and ANOVA, statistical genetics, categorical data analysis, survival analysis, and design and analysis of clinical trials.

M.S., Computer Science, University of Colorado   2005
(Computational Biology emphasis.) Thesis concerned semi-parametric clustering; see "Publications" and "Research" below. Coursework included data mining, algorithms, Bayesian statistics, theory of operating systems, and biochemistry. Awarded the Certificate in Computational Biology.

B.S. *cum laude*, Mathematics, Metropolitan State University of Denver   2001
(Computer Science emphasis.) Coursework included differential equations, discrete mathematics, probability and statistics, numerical analysis, algorithms, databases, general and organic chemistry, general biology, and genetics. Completed requirements for both Computer Science and Biology minors.

A.A.S., Allied Health Sciences, Community College of the Air Force   1997
Coursework included human anatomy and physiology, patient care, and management of medical response.

EMPLOYMENT

Bioinformaticist and Biostatistician, University of Colorado   2013-2019
Research fellow in bioinformatics and integrative physiology at Altitude Research Center, with a focus on human responses and adapatations to acute hypoxia. Bioinformatics lead for Pulmonary Translational Core, Division of Pulmonary Sciences and Critical Care Medicine, providing consulting and services for a wide variety of lung health studies including study planning and trial design, sample size and power calculations, and data analysis.

Training Fellow / Research Assistant, University of Colorado   2007-2013
See "Research" below.

Teaching Assistant, University of Colorado   2009-2010
Assisted in introductory bioinformatics course for first-year students in the Computational Bioscience program. Assisted in annual workshops for biologists to learn bioinformatics skills.

Software Engineer, University of Minnesota 2007
Designed, developed, and maintained clinical and genetic data aggregation, database, and analysis software for Collaborative Study on the Genetics of Asthma. See also "Research" and "Conference Presentations" below.

Research Assistant, University of Minnesota 2005-2007
Responsibilities: see "Research" below.

Teaching Assistant, University of Minnesota 2005-2006
Assisted in classroom-based and online biostatistics courses for health sciences students. Provided individual tutoring, conducted lab sessions, and graded assignments and exams. Developed methods for conveying statistical concepts to students from a wide variety of backgrounds and programs.

Software Engineer, Intelligent Imaging Innovations Inc. 1998-2005
Designed, developed, and maintained corporate database. Served as lead database architect, developer, and administrator. Designed, developed, and maintained commercial end-user software for processing and visualization of microscopic data.

Emergency Medical Technician, Denver Health and other agencies 1997-1998
Performed patient care in pre-hospital, hospital, and clinical environments.

Medical Service Technician ("Medic"), United States Air Force 1989-1997
Performed patient care in pre-hospital, hospital, and clinical environments. Served as Emergency Department shiftleader. Additional duties included supervision and training of junior personnel, clinical IT security, and analysis of medical intelligence.

Infantryman, United States Army 1987-1989

JOURNAL ARTICLES

A. D'Alessandro, T. Nemkov, K. Sun, H. Liu, A. Song, A.A. Monte, A.W. Subudhi, A.T. Lovering, **D. Dvorkin**, C.G. Julian, C.G. Kevil, G.K. Kolluru, S. Shiva, M.T. Gladwin, Y. Xia, K.C. Hansen, and R.C. Roach. AltitudeOmics: Red blood cell metabolic adaptation to high altitude hypoxia. *Journal of Proteome Research*, 15(10):3883-3895, 2016. PMID 27646145, PMCID PMC5512539.

D. Dvorkin, B. Biehs, and K. Kechris. A graphical model method for integrating multiple sources of genome-scale data. *Statistical Applications in Genetics and Molecular Biology*, 12(4):469-87, 2013. PMID 23934610, PMCID 4867227.

S. Wang, **D. Dvorkin**, and Y. Da. SNPEVG: a graphical tool for GWAS graphing with mouse clicks. *BMC Bioinformatics*, 13(1):319, 2012. PMID 23199373, PMCID 3527193.

K. Verspoor, **D. Dvorkin**, K.B. Cohen, and L. Hunter. Ontology quality assurance through analysis of term transformations. *Bioinformatics*, 25(12):i77-84, 2009. PMID 19478020, PMCID 2687949.

L. Ma, H.B. Runesha, **D. Dvorkin**, J.R Garbe, and Y. Da. Parallel and serial computing tools for testing single-locus and epistatic SNP effects of quantitative traits in genome-wide association studies. *BMC Bioinformatics*, 9(1):315, 2008. PMID 18644146, PMCID 2503991.

L. Ma, **D. Dvorkin**, J.R. Garbe, and Y. Da. Genome-wide analysis of single-locus and epistasis single-nucleotide polymorphism effects on anti-cyclic citrullinated peptide as a measure of rheumatoid arthritis. *BMC Proceedings*, 1(Suppl 1):S127, 2007. PMID 18466469, PMCID 2367477.

Y. Mao, N.R. London, L. Ma, **D. Dvorkin**, and Y. Da. Detection of SNP epistasis effects of quantitative traits using an extended Kempthorne model. *Physiological Genomics*, 28(1):46-52, 2006. PMID 16940430.

D. Dvorkin, V. Fadok, and K. Cios. SiMCAL 1 algorithm for analysis of gene expression data related to the phosphatidylserine receptor. *Artificial Intelligence in Medicine*, 35(1-2):49-60, 2005. PMID 16099148.

INDEXED CONFERENCE ABSTRACTS

S. Matson, **D. Dvorkin**, A. Fischer, N. Achtar-Zadeh, P.J. Wolters, and J.S. Lee. Biomarkers in systemic sclerosis associated interstitial lung disease. *American Journal of Respiratory and Critical Care Medicine*, 199:A1432, 2019.

R.C. Roach, **D. Dvorkin**, C.G. Julian, J. Gronewold, N. Bourdillon, J. Bucher, J.E. Elliott et al. Transcriptomic and epigenomic reponses during human adaptation to high-altitude hypoxia. *The FASEB Journal* 29(1 supp): 1051.5, 2016.

C.G. Julian, A. Subudhi, O. Evero, B.D. Pedersen, **D. Dvorkin**, A. Lovering, and R.C. Roach. Epigenetic modification of gene expression during human acclimatization to hypobaric hypoxia. *The FASEB Journal* 28(1 supp): 885.4, 2014.

OTHER CONFERENCE PRESENTATIONS	**D. Dvorkin** and R.C. Roach. Co-expression network hubs in hypoxia and acclimatization to high altitude. Rocky Mountain Bioinformatics Conference, 2014.

D. Dvorkin and R.C. Roach. Co-expression network hubs in hypoxia and acclimatization to high altitude. Rocky Mountain Bioinformatics Conference, 2014.

C.G. Julian, A.W. Subudhi, O. Evero, B.D. Pedersen, **D. Dvorkin**, A.T. Lovering, and R.C. Roach. Genome-wide DNA methylation patterns are altered during human acclimatization to hypobaric hypoxia. Sensing and Signaling of Hypoxia: Interfaces with Biology and Medicine, 2014.

D. Dvorkin and K. Kechris. `lcmix`: an R package for integration of genome-scale data. Rocky Mountain Bioinformatics Conference, 2012.

D. Dvorkin and K. Kechris. Mixture models vs. supervised learning for integrative genomic analysis. Rocky Mountain Bioinformatics Conference, 2011; WNAR/Graybill Conference and NLM Informatics Training Conference, 2012.

D. Dvorkin, B. Biehs, and K. Kechris. Graphical models for integrating multiple sources of genome-scale data. WNAR/IMS Annual Meeting and NLM Informatics Training Conference, 2011.

D. Dvorkin, B. Biehs, and K. Kechris. Identifying genes in the Drosophila Hh pathway by integrating TF binding and gene expression data. Rocky Mountain Bioinformatics Conference, 2010.

D. Dvorkin and K. Kechris. Development of methods for integrating diverse sources of genome-scale data. Rocky Mountain Bioinformatics Conference, 2009.

K. Verspoor, **D. Dvorkin**, K.B. Cohen, and L. Hunter. Ontology quality assurance through analysis of term transformations. Intelligent Systems for Molecular Biology and European Conference on Computational Biology, 2009.

D. Dvorkin, M.B. Miller, M.N. Blumenthal, S.A. Kelleher, and J.P. Raasch. Identification of significantly different from expected changes in pulmonary function test values over time. International Conference of the American Thoracic Society, 2008.

L. Ma, **D. Dvorkin**, and Y. Da. Large scale SNP epistasis detection of complex traitsusing pairwise epistasis tests. Annual Meeting of the International Genetic Epidemiology Society, 2006.

S. Basu, T.L. Bergemann, **D. Dvorkin**, and M. Seth. The analysis of quantitative traits correlated with rheumatoid arthritis. Genetic Analysis Workshop, 2006.

RESEARCH	Prediction of aspiration in ventilator patients	2018-2019

Prediction of aspiration in ventilator patients   2018-2019
Development of clinical decision support trees to predict aspiration from bedside exam data in patients recently removed from ventilators.

Biomarkers in systemic sclerosis associated interstitial lung disease   2018-2019
Biomarker-based outcome prediction of morbidity and mortality in patients suffering from systemic sclerosis. See "Indexed Conference Abstracts" above.

Integrative physiology and bioinformatics of altitude and hypoxia   2013-2019
AltitudeOmics: predictive models for gene expression, epigenomic regulation, and control of physiological traits in acute exposure and long-term acclimatization to hypoxia induced by exposure to high altitude. See "Publications" above.

Integrating diverse sources of genome-scale data   2009-2013 (dissertation research)
Development and evaluation of graphical mixture models to integrate multiple data sources such as expression, binding, and DNA sequence conservation to present a more complete picture of gene regulation and expression, and identify genes and *cis*-regulatory regions which play key roles within certain pathways. See "Publications" above.

Detection of sequence motifs correlated with expression   2008
Improved implementation of c-REDUCE software for discovery of gene sequence motifs which show significant correlation with gene expression. Algorithm analysis, code refactoring for improved performance, and design and development of new end-user features.

Measures for ontology quality evaluation   2007-2009
Design and development of algorithms, significance tests, and applications for measures of quality in the Gene Ontology (GO) and other ontologies. See "Publications" above.

Collaborative Study on the Genetics of Asthma (CSGA)   2007-2009
Design and development of procedures and applications for identification and visualization of significant changes in pulmonary function test values. See "Conference Presentations" above.

Prediction of gene coexpression using the Gene Ontology   2007 (M.S. project research)
Design and development of procedures for using organism-specific GO annotations to calculate similarities between genes and predict coexpression as a tool to improve analysis of microarray experiments.

Transcription factor target detection in *E. coli*   2006
Analysis of sequence and expression data for detection of transcription factor targets. Integration of sequence analysis tools such as AlignACE/ScanACE and MEME/MAST into end-user applications.

Genetic Analysis Workshop (GAW) 15   2006
Trait selection, data management, inheritance modeling, and data visualization for SNP marker analysis of quantitative traits in rheumatoid arthritis. See "Conference Presentations" above.

Epistasis effects in quantitative traits in cattle   2006-2012
Design and development of end-user software for visualization of genetic aspects of birth weight, milk production, and fertility. See "Publications" above.

Feasibility of Retinoic Acid Treatment in Emphysema (FORTE)   2005-2006
Clinical trials data management, analysis, and reporting. Aggregation of data from multiple sources in multiple formats, database management, and data mining.

The SiMCAL family of algorithms for analysis of microarray data   2003-2005 (thesis research)
Design and development of novel clustering algorithms and visualization tools for analysis of microarray data in the action of the phosphatidylserine receptor in inflammatory response, and in expression changes in cystic fibrosis patients. See "Publications" above.

HONORS AND AWARDS		
	ORISE Postgraduate Research Fellowship	2018-2019
	NIH/NLM Training Fellowship T15 LM009451 Computational Bioscience, University of Colorado	2008-2013
	University Fellowship, The Ohio State University (declined)	2007
	NIH Training Fellowship, Biostatistics in Psychiatric Research, U. of Pittsburgh (declined)	2007
	B.S. *cum laude*, Metropolitan State College of Denver	2001
	Honor Graduate, USAF Medical Service Training Course	1990

SOFTWARE AND DATABASES

Programming languages: considerable experience in R, Python, and PHP. Moderate experience in C, C++, and Perl. Some experience in JavaScript and Java. Released open-source software in R (`http://r-forge.r-project.org/projects/lcmix/`) and Python (`http://sourceforge.net/projects/dvpython/`).

Database management systems: considerable experience in MySQL and SQLite. Some experience in Oracle and PostgreSQL.

Other: familiar with Unix (macOS, Linux, BSD, Solaris) and Windows environments. Considerable experience in HTML. Moderate experience in SAS. Some experience in Linux and BSD system and network administration.