William Stafford Noble*

Department of Genome Sciences
University of Washington
Foege Building, S340A
3720 15th Ave NE, Box 357730
Seattle, WA 98195
Tel: (206) 543–8930   Fax: (206) 685–7301
wnoble@uw.edu
http://noble.washington.edu/∼noble
Pronouns: he/him

Education

1991	B.S. with honors and distinction in Symbolic Systems, concentrating in Philosophical Foundations, Stanford University. Prof. Fred Dretske, advisor.
1996	M.S. in Computer Science, University of California, San Diego.
1998	Ph.D. in Computer Science and Cognitive Science, University of California, San Diego. Prof. Charles Elkan, advisor.

Employment and Professional Appointments

1989	Center for Strategic Technology and Research, Andersen Consulting, Chicago, IL (internship).
1990–91	SRI International, Menlo Park, CA.
1991–93	United States Peace Corps Volunteer, Lesotho, southern Africa.
1994	Entropic Research Laboratory, Palo Alto, CA.
1998–99	Postdoctoral fellow, University of California, Santa Cruz. Prof. David Haussler, advisor.
1999–02	Assistant Professor, Department of Computer Science, Columbia University, with joint appointment at the Columbia Genome Center.
2002–06	Assistant Professor, Department of Genome Sciences, University of Washington
2002–06	Adjunct Assistant Professor, Department of Computer Science and Engineering, University of Washington
2005–06	Adjunct Assistant Professor, Department of Medicine, University of Washington
2006–11	Associate Professor, Department of Genome Sciences, University of Washington
2006–11	Adjunct Associate Professor, Department of Computer Science and Engineering, University of Washington
2006–11	Adjunct Associate Professor, Department of Medicine, University of Washington
2013–20	Director, Computational Molecular Biology Program, University of Washington
2011–	Professor, Department of Genome Sciences, University of Washington
2011–	Adjunct Professor, Department of Computer Science and Engineering, University of Washington
2011–	Adjunct Professor, Department of Medicine, University of Washington
2012–	Adjunct Professor, Department of Biomedical Informatics and Medical Education, University of Washington
2014–	Senior Data Science Fellow, University of Washington eScience Institute
2020–21	Interim Chair, Department of Genome Sciences, University of Washington

---

*Formerly William Noble Grundy: see http://noble.gs.washington.edu/˜noble/name-change.html

Awards

1987	National Merit Scholar
1987	David Starr Jordan Scholar, Stanford University
1991	Phi Beta Kappa, Stanford University
1994–97	Fellow, National Defense Science and Engineering Graduate Fellowship Program
1998–99	Fellow, Alfred P. Sloan Foundation and U.S. Department of Energy Postdoctoral Fellowships in Computational Molecular Biology
2000–02	Pharmaceutical Research and Manufacturers of America Foundation Faculty Development Award in Bioinformatics
2001–06	National Science Foundation CAREER Award
2001–05	Research Fellow, Alfred P. Sloan Foundation
2010	University of Washington Postdoc Mentor of the Year
2013	Fulbright Specialist Grant, African Institute for Mathematical Sciences, Muizenberg, South Africa
2015–6, 18–23	Clarivate Analytics list of "Highly cited researchers"
2017	Fellow, International Society for Computational Biology
2018–19	Chair, BDMA review panel, National Institutes of Health
2019	Innovator Award, International Society for Computational Biology
2021	Team Science Award, ENCODE Consortium, National Institutes of Health
2023	Fellow, American Institute for Medical and Biological Engineering

Professional Activities

Board memberships Member, Board of Directors, International Society for Computational Biology, 2008–2011.

Editorial boards

- Guest co-editor, Special issue on Machine Learning for Bioinformatics, *IEEE Transactions on Computational Biology and Bioinformatics*, 2004.

- *Journal of Bioinformatics and Computational Biology*, 2004–2020.

- *IEEE Transactions on Computational Biology and Bioinformatics*, 2005–2020.

- *Quantitative Biology*, 2012–2020.

- *Rapid Communications in Mass Spectrometry*, Editorial Advisory Board, 2018–2021.

- *PLOS Computational Biology*, Associate Editor, 2008–2012; Deputy Editor, 2012–present.

- *BMC Bioinformatics*, Associate Editor, 2018–present.

Scientific advisory boards

- X-Mine, Inc., Hayward, CA, 2000–2002.

- Bioinformatics of Mammalian Gene Expression project, Canada's Michael Smith Genome Centre, Vancouver, BC, Canada, 2004–2006.

- Center for Functional Genomics and HCV-Associated Liver Disease, University of Washington, 2006–2007.

- National Center for Systems Biology, Institute for Systems Biology, Seattle, WA, 2006–2009.

- DNAnexus, Mountain View, CA, 2012–2014.

Review panels   National Institutes of Health:

- Special Bioinformatics Study Section, March 12, 2003

- Special Bioinformatics Study Section, June 30, 2004

- Bioengineering Research Partnership Study Section, 13 Dec 2004

- Special Bioinformatics Study Section, March 17–18, 2005

- Bioengineering Sciences and Technologies special emphasis panel on "Microarray data analysis, data integration and data exploration," August 5, 2005

- National Cancer Institute special emphasis panel on "Advanced proteomic platforms and computation science for the NCI clinical proteomic technologies initiative," June 26–27, 2006

- National Cancer Institute special emphasis panel on "Advanced Genomic Data Analysis and Visualization Methods for TCGA Data," April 2, 2008

- Topics in Biomedical Engineering, February 14, 2011.

- GCAT special emphasis panel, July 5, 2011.

- HLBP1 Workgroup, January 11, 2012.

- NIDDK—Collaborative Interdisciplinary Team Science Research, February 19, 2013.

- Biodata Management and Analysis, Jan 2007, Oct 2008, Feb 2009, Oct 2011, Feb 2012, Oct 2012, Jun 2013

- Member, Biodata Management and Analysis, July, 2014–Sep, 2017.

- Chair, Biodata Management and Analysis, October, 2017–Jun, 2019.

- Topics in Bacterial Pathogenesis, March 18, 2020.

- Member Conflicts — Bioengineering, Biodata, and Biomodeling Technologies, July 29, 2022.

- Program for Inclusion and Diversity Among Individuals Engaged in Health Related Research (PRIDE) R25 and U24, July 24, 2023.

Panelist, National Science Foundation

- Information Technology Research at the Intersection of Biology and Informatics, 2001.

- Advances in Biological Informatics, Division of Biological Infrastructure, 2015.

Program committee memberships   Machine learning and artificial intelligence conferences:

- American Association for Artificial Intelligence 1998, 2016–2020 (Senior PC member for 2018).

- Knowledge Discovery in Databases 2000, 2003

- Biological Knowledge Discovery in Databases 2002, 2013

- Conference on Learning Theory 2003

- International Conference on Machine Learning 2004, 2008, 2013–2016, 2019–2023.

- International Joint Conference on Artifical Intelligence 2009

- Neural Information Processing Systems, 2003–2016, 2018–2023

- Neural Information Processing Systems Workshop on Machine Learning Open Source Software 2006, 2008

- Neural Information Processing Systems Workshop on Mining and Learning with Graphs 2008, 2009

- Neural Information Processing Systems Workshop on Machine Learning in Computational Biology 2005–2017, 2019–2023

- Machine Learning in Systems Biology 2009

- Uncertainty in Artificial Intelligence 2013–2016

- AISTATS 2017–2021

- International Conference on Learning Representations 2019–2020, 2022–2023

Bioinformatics conferences:

- Intelligent Systems for Molecular Biology 2002–2009, 2011–2016, 2018–2024.

- Research in Computational Biology (RECOMB) 2004, 2007–2015, 2017–2024

- RECOMB Computational Proteomics Satellite 2007, 2010–2012

- RECOMB Massively Parallel Sequencing Satellite, 2012, 2014

- RECOMB Highlights Committee 2023.

- Genome Informatics Workshop 2005–2009

- European Conference on Computational Biology 2005

- Bioinformatics Research and Development (BIRD) 2007

- Computational Systems Biology 2009

- ACM Conference on Bioinformatics, Computational Biology and Biomedicine (ACM-BCB) 2011, 2015–2016, 2020-2022, 2024.

- ASE/IEEE International Conference on BioMedical Computing 2012–2013

- BioKDD 2013, 2016, 2019–2020, 2022

- ISCB Africa ASBCB 2013

- High Throughput Sequencing Conference (HitSeq) 2014–2015

- Workshop on Algorithms in Bioinformatics 2015–2016, 2019–2021.

- Workshop on Machine Learning in Systems Biology 2016

- International Conference on Algorithms for Computational Biology 2017–2019

- IEEE International Conference on BioInformatics and BioEngineering (BIBE) 2018

- RECOMB/ISCB Conference on Regulatory and Systems Genomics with DREAM Challenges 2020

Area chair, Bioinformatics and Kernel Methods, Twenty-first Annual Conference on Neural Information Processing Systems, Whistler, BC, Dec 3–8, 2007.

Co-chair, Machine Learning in Computational Biology Workshop, Twentieth Annual Conference on Neural Information Processing Systems, Whistler, BC, 2005–2008.

Member, College of CSR Reviewers, National Institutes of Health. March, 2010–present.

Area chair, Bioinformatics and Kernel Methods, Twenty-fourth Annual Conference on Neural Information Processing Systems, Whistler, BC, Dec 6–11, 2010.

Co-chair, Workshop on the Structure and Function of Chromatin and Chromosomes, Pacific Symposium on Biocomputing, The Big Island of Hawaii, January 3–7, 2012.

Co-chair, Special Session on Computational Methods for Elucidating Nuclear Structure and Dynamics, Intelligent Systems for Molecular Biology Conference, Long Beach, CA, USA, July 15–17, 2012.

Co-chair, Epigenomics session, Pacific Symposium on Biocomputing, The Big Island of Hawaii, January 3–7, 2013.

Co-chair, Mass Spectrometry & Proteomics Track, Intelligent Systems for Molecular Biology Conference, July 11-15, 2014.

Co-chair, Bioinformatics Group, American Society for Mass Spectrometry Annual Conferences, 2019 and 2020.

Co-chair, CompMS Community of Special Interest for ISCB, 2018–2020.

Journal referee

	2015	2016	2017	2018	2019	2020	2021	2022
Analytical Chemistry	2							
BMC Bioinformatics							1	
Bioinformatics	2	1	3	3	1			1
Chromosome Research		1						
Cell Systems				1				
Communications Biology						1		
eLife		1	1					
Epigenomics & Chromatin					1			
FEBS Letters	1							
Genome Biology								2
Genome Research	4	4	1	2	2	1		1
Genes					1			
G3				1				
Journal of Proteome Research	1	2	1	3	1		1	3
Molecular & Cellular Proteomics	1							
Molecular Systems Biology							1	
Nucleic Acids Research		1						
Nature	1	1	4	1	4			1
Nature Biotechnology	1	1	1	1	1	1	2	
Nature Communications	1	2	1	2		1	1	
Nature Evolution & Ecology							1	
Nature Genetics				1	1	2	1	
Nature Materials				1				
Nature Methods	2	1		1		1	3	2
Nature Reviews Genetics		1						
Nature Structural Biology	1							
Nature Structural Molecular Biology		1						
Nucleic Acids Research		1				1		
PLOS One		1						
PLOS Computational Biology				1	1			1
PNAS		2	1					
Proteomics		1						
Science			1	5		1		
Scientific Reports					1			1
Tenure & promotion review	5	7	4	7	3	8	1	5
Total	22	28	18	28	17	18	11	17

Committee memberships   Member, Public Affairs and Policies Committee, International Society for Computational Biology, 2003–present.

Audit Committee, International Society for Computational Biology, 2012.

Computational biology search committee, Fred Hutchinson Cancer Research Center, 2010–2011.

Faculty search committee, Department of Biochemistry, computational biology and protein design, 2013–2014, 2014–2015.

Refereed publications

Underlined names are lab members.

(1) **WN Grundy**, TL Bailey and CP Elkan. "ParaMEME: a parallel implementation and a web interface for a DNA and protein motif discovery tool." *Computer Applications in the Biosciences.* 12(4):303-310, 1996.

(2) J Batali and **WN Grundy**. "Modeling the evolution of motivation." *Evolutionary Computation.* 4(3):235-270, 1997.

(3) **WN Grundy**, TL Bailey, CP Elkan and ME Baker. "Hidden Markov model analysis of motifs in steroid dehydrogenases and their homologs." *Biochemical and Biophysical Research Communications.* 231(3):760-766, 1997.

(4) **WN Grundy**, TL Bailey, CP Elkan and ME Baker. "Meta-MEME: Motif-based hidden Markov models of protein families." *Computer Applications in the Biosciences.* 13(4):397-406, 1997.

(5) ME Baker, **WN Grundy** and CP Elkan. "Spinach CSP41, an mRNA-binding protein and ribonuclease, is homologous to nucleotide-sugar epimerases and hydroxysteroid dehydrogenases." *Biochemical and Biophysical Research Communications.* 248(2):250-254, 1998.

(6) **WN Grundy**. "Homology detection via Family Pairwise Search." *Journal of Computational Biology.* 5(3):479-492, 1998.

(7) **WN Grundy**. "Family-based homology detection via pairwise sequence comparison." *Proceedings of the Second Annual International Conference on Computational Molecular Biology*, March 22-25, 1998. pp. 94-100.

(8) ME Baker, **WN Grundy** and CP Elkan. "A common ancestor for a subunit in the mitochondrial proton-translocating NADH:ubiquinone oxidoreductase (complex I) and short-chain dehydrogenases/reductases." *Cellular and Molecular Life Sciences.* 55(3):450-455, 1999.

(9) **WN Grundy** and TL Bailey. "Family Pairwise Search with embedded motif models." *Bioinformatics.* 15(6):463-470, 1999.

(10) **WN Grundy** and GJP Naylor. "Phylogenetic inference from conserved sites alignments." *Journal of Experimental Zoology.* 285(2):128-139, 1999.

(11) TL Bailey and **WN Grundy**. "Classifying proteins by family using the product of correlated p-values." *Proceedings of the Third International Conference on Computational Molecular Biology*, April 11-14, 1999. pp. 10-14.

(12) MPS Brown, **WN Grundy**, D Lin, N Cristianini, C Sugnet, TS Furey, M Ares, Jr. and D Haussler. "Knowledge-based analysis of microarray gene expression data by using support vector machines." *Proceedings of the National Academy of Science.* 97(1):262-267, 2000.

(13) E Eskin, **WN Grundy** and Y Singer. "Protein family classification using sparse markov transducers." *Proceedings of the Eighth International Conference on Intelligent Systems for Molecular Biology*, August 20-23, 2000. pp. 134–145.

(14) P Pavlidis, TS Furey, M Liberto, D Haussler and **WN Grundy**. "Promoter region-based classification of genes." *Proceedings of the Pacific Symposium on Biocomputing*, January 3-7, 2001. pp. 151-163.

(15) P Pavlidis, J Weston, J Cai and **WN Grundy**. "Gene functional classification from heterogeneous data." *Proceedings of the Fifth International Conference on Computational Molecular Biology*, April 21-24, 2001. pp. 242-248.

(16) E Eskin, **WN Grundy** and Y Singer. "Using mixtures of common ancestors for estimating the probabilities of discrete events in biological sequences." *Bioinformatics (Proceedings of the ISMB).* 17(Suppl 1):S64–73, 2001.

(17) P Pavlidis, C Tang and **WS Noble**. "Classification of genes using probabilistic models of microarray expression profiles." *Proceedings of BIOKDD 2001: Workshop on Data Mining in Bioinformatics.* August 26, 2001. pp. 15-21.

(18) RA Muhle, <u>P Pavlidis</u>, **WN Grundy** and E Hirsch. "A high throughput study of gene expression in preterm labor using a subtractive microarray approach." *American Journal of Obstetrics and Gynecology.* 185(3):716-24, 2001.

(19) <u>P Pavlidis</u> and **WS Noble**. "Analysis of strain and regional variation in gene expression in mouse brain." *Genome Biology.* 2(10): research0042.1-0042.15, 2001.

(20) <u>P Pavlidis</u>, J Weston, <u>J Cai</u> and **WS Noble**. "Learning gene functional classifications from multiple data types." *Journal of Computational Biology.* 9(2):401-411, 2002.

(21) CS Leslie, <u>E Eskin</u> and **WS Noble**. "The spectrum kernel: An SVM-string kernel for protein classification." *Proceedings of the Pacific Symposium on Biocomputing*, January 2-7, 2002. pp. 564–575.

(22) <u>P Pavlidis</u>, <u>DP Lewis</u> and **WS Noble**. "Exploring gene expression data with class scores." *Proceedings of the Pacific Symposium on Biocomputing*, January 2-7, 2002. pp. 474–485.

(23) <u>E Eskin</u>, **WS Noble** and Y Singer. "Using substitution matrices to estimate probability distributions for biological sequences." *Journal of Computational Biology.* 9(6):775-791, 2002.

(24) <u>L Liao</u> and **WS Noble**. "Combining pairwise sequence similarity and support vector machines for remote protein homology detection." *Proceedings of the Sixth International Conference on Computational Molecular Biology*, April 18-21, 2002. pp. 225–232.

(25) B Schölkopf, J Weston, <u>E Eskin</u>, CS Leslie and **WS Noble**. "A kernel approach for learning from almost orthogonal patterns." *Proceedings of the 13th European Conference on Machine Learning*, August 19-23, 2002. pp. 511-528.

(26) B Schölkopf, J Weston, E Eskin, CS Leslie and **WS Noble**. "Dealing with large diagonals in kernel matrices." *Annals of the Institute on Statistical Mathematics*, 55(2):391–408, 2003.

(27) <u>P Pavlidis</u> and **WS Noble**. "Matrix2png: A utility for visualizing matrix data." *Bioinformatics.* 19(2):295-296, 2003.

(28) NH Segal, <u>P Pavlidis</u>, **WS Noble**, CR Antonescu, A Viale, UV Wesley, K Busam, H Gallardo, D DeSantis, MF Brennan, C Cordon-Cardo, JD Wolchok and AN Houghton. "Classification of clear cell sarcoma as melanoma of soft parts by genomic profiling." *Journal of Clinical Oncology.* 21:1775–1781, 2003.

(29) DC Anderson, W Li, DG Payan and **WS Noble**. "A new algorithm for the evaluation of shotgun peptide sequencing in proteomics: support vector machine classification of peptide MS/MS spectra and SEQUEST scores" *Journal of Proteome Research.* 2(2):137–146, 2003.

(30) NH Segal, <u>P Pavlidis</u>, CR Antonescu, RG Maki, **WS Noble**, JM Woodruff, JJ Lewis, MF Brennan, AN Houghton and C Cordon-Cardo. "Classification and subtype prediction of soft tissue sarcoma by functional genomics and support vector machine analysis." *American Journal of Pathology.* 169:691-700, 2003.

(31) T Gururaja, W Li, **WS Noble**, DG Payan and DC Anderson. "Multiple functional categories of proteins identified in an *in vitro* cellular ubiquitin affinity extract using shotgun peptide sequencing." *Journal of Proteome Research.* 2:383–393, 2003.

(32) <u>P Pavlidis</u>, <u>Q Li</u> and **WS Noble**. "The effect of replication on gene expression microarray experiments." *Bioinformatics.* 19(13):1620-1627, 2003.

(33) <u>J Qin</u>, <u>DP Lewis</u> and **WS Noble**. "Kernel hierarchical clustering of microarray gene expression data." *Bioinformatics.* 19:2097-2014, 2003.

(34) <u>E Eskin</u>, **WS Noble** and Y Singer. "Protein family classification using sparse Markov transducers." *Journal of Computational Biology.* 10(2):187–213, 2003.

(35) <u>L Liao</u> and **WS Noble**. "Combining pairwise sequence similarity and support vector machines for detecting remote protein evolutionary and structural relationships." *Journal of Computational Biology.* 10(6):857–868, 2003.

(36) CS Leslie, <u>E Eskin</u>, J Weston and **WS Noble**. "Mismatch string kernels for SVM protein classification." *Advances in Neural Information Processing Systems 15*, 2003.

(37) TL Bailey and **WS Noble**. "Searching for statistically significant regulatory modules." *Bioinformatics (Proceedings of the European Conference on Computational Biology).* 19(Suppl. 2):ii16–ii25, 2003.

(38) SM Gomez, **WS Noble** and A Rzhetsky. "Learning to predict protein-protein interactions from protein sequences." *Bioinformatics (Proceedings of the Georgia Tech International Conference on Bioinformatics).* 19:1875–1881, 2003.

(39) J Weston, A Eliseeff, D Zhou, CS Leslie and **WS Noble.** "Protein ranking: From local to global structure in the protein similarity network." *Proceedings of the National Academy of Science.* 101(17):6559–6563, 2004.

(40) CS Leslie, <u>E Eskin</u>, A Cohen, J Weston and **WS Noble**. "Mismatch string kernels for discriminative protein classification." *Bioinformatics.* 20(4):467–476, 2004.

(41) <u>P Pavlidis</u>, I Wapinski and **WS Noble**. "Support vector machine classification on the web." *Bioinformatics.* 20(4):586–587, 2004.

(42) <u>W Wu</u> and **WS Noble**. "Genomic data visualization on the web." *Bioinformatics.* 20(11):1804–1805, 2004.

(43) K Tsuda and **WS Noble**. "Learning kernels from biological networks by maximizing entropy." *Bioinformatics (Proceedings of the Intelligent Systems for Molecular Biology Conference).* 20(Suppl. 1):i326–i333, 2004.

(44) J Weston, CS Leslie, D Zhou and **WS Noble**. "Semi-supervised protein classification using cluster kernels." *Advances in Neural Information Processing Systems 16*, 2004. pp. 595–602.

(45) GRG Lanckriet, M Deng, N Cristianini, MI Jordan and **WS Noble**. "Kernel-based data fusion and its application to protein function prediction in yeast." *Proceedings of the Pacific Symposium on Biocomputing*, January 3-8, 2004. pp. 300-311.

(46) GRG Lanckriet, T De Bie, N Cristianini, MI Jordan and **WS Noble**. "A statistical framework for genomic data fusion." *Bioinformatics.* 20(16):2626-2635, 2004.

(47) H Lu, W Li, **WS Noble**, DG Payan and DC Anderson. "Riboproteomics of the hepatitis C virus internal ribosomal entry site." *Journal of Proteome Research* 3(5):949–57, 2004.

(48) E Feingold, PJ Good, . . . , **WS Noble**, . . . , FS Collins. "The ENCODE (ENCyclopedia Of DNA Elements) Project." *Science.* 306:636–640, 2004.

(49) M Tompa, N Li, TL Bailey, GM Church, B De Moor, E Eskin, AV Favorov, MC Frith, Y Fu, WJ Kent, VJ Makeev, AA Mironov, **WS Noble**, G Pavesi, G Pesole, M Régnier, N Simonis, S Sinha, G Thijs, J van Helden, M Vandenbogaert, Z Weng, C Workman, C Ye and Z Zhu. "Assessing computational tools for the discovery of transcription factor binding sites." *Nature Biotechnology.* 23(1):137–144, 2005.

(50) <u>A Ben-Hur</u> and **WS Noble**. "Kernel methods for predicting protein-protein interactions." *Bioinformatics (Proceedings of the Intelligent Systems for Molecular Biology Conference).* 21(Suppl 1):i38–i46, 2005.

(51) **WS Noble**, MS Kuehn, <u>RE Thurman</u>, R Humbert, JC Wallace, M Yu, M Hawrylycz and JA Stamatoyannopoulos. "Predicting the *in vivo* signature of human gene regulatory sequences." *Bioinformatics (Proceedings of the Intelligent Systems for Molecular Biology Conference).* 21(Suppl 1):i338–i343, 2005.

(52) J Weston, CS Leslie, E Ie, D Zhou, A Eliseeff and **WS Noble**. "Semi-supervised protein classification using cluster kernels." *Bioinformatics.* 21(15):3241–3247, 2005.

(53) W Sheffler, E Upfal, J Sedivy and **WS Noble**. "A learned comparative expression measure for Affymetrix GeneChip DNA microarrays." *Proceedings of the Computational Systems Bioinformatics Conference*, August 8-11, 2005, Stanford, CA. pp. 144-154.

(54) T Mann, R Humbert, JA Stamatoyannopoulos and **WS Noble**. "Automated validation of polymerase chain reactions using amplicon melting curves." *Proceedings of the Computational Systems Bioinformatics Conference*, August 8-11, 2005, Stanford, CA. pp. 377–385.

(55) AA Klammer, CW Wu, MJ MacCoss, **WS Noble**. "Peptide charge state determination for low-resolution tandem mass spectra." *Proceedings of the Computational Systems Bioinformatics Conference*, August 8-11, 2005, Stanford, CA. pp. 175–185.

(56) E Ie, J Weston, **WS Noble** and CS Leslie. "Adaptive codes for multi-class protein classification." *Proceedings of the International Conference on Machine Learning*, August 7-11, 2005, Bonn, Germany.

(57) R Kuang, J Weston, **WS Noble** and CS Leslie. "Motif-based protein ranking by network propagation." *Bioinformatics.* 21(19):3711–3718, 2005.

(58) JP Miller, RS Lo, A Ben-Hur, C Desmarais, I Stagljar, **WS Noble** and S Fields. "Large-scale identification of yeast integral membrane protein interactions." *Proceedings of the National Academy of Science.* 102(34):12123–12128, 2005.

(59) JP Vert, RE Thurman and **WS Noble**. "Kernels for gene regulatory regions." *Advances in Neural Information Processing Systems 19.* 2006.

(60) J Weston, R Kuang, CS Leslie and **WS Noble**. "Protein ranking by semi-supervised network propagation." *BMC Bioinformatics.* 7(Suppl 1):S10, 2006.

(61) A Ben-Hur and **WS Noble**. "Choosing negative examples for the prediction of protein-protein interactions." *BMC Bioinformatics.* 7(Suppl 1):S2, 2006.

(62) DP Lewis, T Jebara and **WS Noble**. "Nonstationary kernel combination." *Proceedings of the International Conference on Machine Learning*, June 25-29, 2006, Pittsburgh, PA.

(63) T Mann and **WS Noble**. "Efficient identification of DNA binding partners in a sequence database." *Bioinformatics (Proceedings of the Intelligent Systems for Molecular Biology Conference).* 22(14):e350–e358, 2006.

(64) PJ Sabo, MS Kuehn, RE Thurman, B Johnson, EM Johnson, H Cao, M Yu, J Goldy, E Rosenzweigh, J Goldy, A Haydock, M Weaver, A Shafer, K Lee, F Neri, R Humbert, MA Singer, TA Richmond, MO Dorschner, M McArthur, M Hawrylycz, RD Green, PA Navas, **WS Noble** and JA Stamatoyannopoulos. "Genome-scale mapping of DNase I sensitivity *in vivo* using tiling DNA microarrays." *Nature Methods.* 3(7):511–518, 2006.

(65) BE Frewen, GE Merrihew, **WS Noble** and MJ MacCoss. "Analysis of peptide MS/MS spectra from large-scale proteomics experiments using spectrum libraries." *Analytical Chemistry.* 78(16):5678–5684, 2006.

(66) T Pramila, W Wu, **WS Noble** and LL Breeden. "The Forkhead transcription factor Hcm1 regulates chromosome segregation genes and fills the S phase gap in the transcriptional circuitry of the cell cycle." *Genes and Development.* 20(16):2266–2278, 2006.

(67) T Mann, R Humbert, JA Stamatoyannopoulos and **WS Noble**. "Automated validation of polymerase chain reactions using amplicon melting curves." *Journal of Bioinformatics and Computational Biology.* 22(14):350–358, 2006.

(68) DP Lewis, T Jebara and **WS Noble**. "Support vector machine learning from heterogeneous data: an empirical analysis using protein sequence and structure." *Bioinformatics.* 22(22):2753–2760, 2006.

(69) S Gupta, JA Stamatoyannopoulos, TL Bailey and **WS Noble**. "Quantifying similarity between motifs." *Genome Biology.* 8:R24, 2007.

(70) AA Klammer, X Yi, MJ MacCoss, and **WS Noble**. "Improving tandem mass spectrum identification using peptide retntion time prediction across diverse chromatography conditions." *Analytical Chemistry.* 79(160):6111-6118, 2007.

(71) J Qiu, M Hue, A Ben-Hur, JP Vert and **WS Noble**. "A structural alignment kernel for protein structures." *Bioinformatics.* 23(9):1090-1098, 2007.

(72) AA Klammer, X Yi, MJ MacCoss and **WS Noble**. "Peptide retention time prediction yields improved tandem mass spectrum identification for diverse chromatography conditions." *Proceedings of the International on Research in Computational Biology (RECOMB).* April 21–25, 2007. pp. 459–472.

(73) The ENCODE Project Consortium. "Identification and analysis of functional elements in 1% of the human genome by the ENCODE project." *Nature.* 447:779–816, 2007.

(74) RE Thurman, N Day, **WS Noble** and JA Stamatoyannopoulos. "Identification of higher-order functional domains in the human ENCODE regions." *Genome Research.* 17(6):917–927, 2007.

(75) JH Dennis, H-Y Fan, SM Reynolds, G Yuan, J Meldrim, DJ Richter, DG Peterson, OJ Rando, **WS Noble** and RE Kingston. "Independent and complementary methods for large-scale structural analysis of mammalian chromatin." *Genome Research.* 17(6):928-939, 2007.

(76) I Melvin, E Ie, J Weston, **WS Noble** and CS Leslie. "Multi-class protein classification using adaptive codes." *Journal of Machine Learning Research.* 8:1557–1581, 2007.

(77) N Day, A Hemmaplardh, RE Thurman, JA Stamatoyannopoulos and **WS Noble**. "Unsupervised segmentation of continuous genomic data." *Bioinformatics.* 23(11):1424–1426, 2007.

(78) S Asthana, **WS Noble**, G Kryukov, CE Grant, S Sunyaev and JA Stamatoyannopoulos. "Widely distributed non-coding selection in the human genome." *Proceedings of the National Academy of Science.* 104(30):12410–12415, 2007.

(79) H Peckham, RE Thurman, Y Fu, JA Stamatoyannopoulos, **WS Noble**, K Struhl and Z Weng. "Nucleosome positioning signals in genomic DNA." *Genome Research.* 17(8):1170–1177, 2007.

(80) AA Klammer, X Yi, MJ MacCoss and **WS Noble**. "Peptide retention time prediction yields improved tandem mass spectrum identification for diverse chromatography conditions." *Analytical Chemistry.* 79(16):6111–6118, 2007.

(81) L Käll, J Canterbury, J Weston, **WS Noble** and MJ MacCoss. "Semi-supervised learning for peptide identification from shotgun proteomics datasets." *Nature Methods.* 4(11):923–925, 2007.

(82) I Melvin, E Ie, R Kuang, J Weston, **WS Noble** and CS Leslie. "SVM-fold: a tool for discriminative multi-class protein fold and superfamily recognition." *BMC Bioinformatics.* 8(Suppl 4):S2, 2007.

(83) JP Vert, J Qiu and **WS Noble**. "A new pairwise kernel for biological network inference with support vector machines." *BMC Bioinformatics.* 8(Suppl. 10):S8, 2007.

(84) J Qiu, W Sheffler, D Baker and **WS Noble**. "Ranking predicted protein structures with support vector regression." *Proteins: Structure, Function, and Bioinformatics.* 71(3):1175–1182, 2007.

(85) RE Thurman, **WS Noble** and JA Stamatoyannopoulos. "Multi-scale correlations in continuous genomic data." *Proceedings of the Pacific Symposium on Biocomputing.* January 4–8, 2008, pp. 201-215.

(86) L Käll, JD Storey, MJ MacCoss and **WS Noble**. "Assigning significance to peptides identified by tandem mass spectrometry using decoy databases." *Journal of Proteome Research.* 7(1):29–34, 2008.

(87) J Qiu and **WS Noble**. "Predicting co-complexed protein pairs from heterogeneous data." *PLOS Computational Biology*. 4(4):e1000054, 2008.

(88) L Peña-Castillo, M Tasan, CL Myers, H Lee, T Joshi, C Zhang, Y Guan, M Leone, A Pagnani, WK Kim, C Krumpelman, W Tian, G Obozinski, Y Qi, S Mostafavi, GN Lin, G Berriz, F Gibbons, GRG Lanckriet, J Qiu, CE Grant, Z Barutcuoglu, DP Hill, D Warde-Farely, C Grouios, D Ray, JA Blake, M Deng, MI Jordan, **WS Noble**, QD Morris, J Klein-Seetharaman, Z Bar-Joseph, T Chen, F Sun, OG Troyanskaya, EM Marcotte, D Xu, TR Hughes and FP Roth. "A critical assessment of *M. musculus* gene function prediction using integrated genomic evidence." *Genome Biology*. 9(Suppl 1):S2, 2008.

(89) G Obozinski, GRG Lanckriet, CE Grant, MI Jordan and **WS Noble**. "Consistent probabilistic outputs for protein function prediction." *Genome Biology*. 9(Suppl 1):6, 2008.

(90) AA Klammer, SM Reynolds, JA Bilmes, MJ MacCoss and **WS Noble**. "Modeling peptide fragmentation with dynamic Bayesian networks for peptide identification." *Bioinformatics (Proceedings of the Intelligent Systems for Molecular Biology Conference)*. 24(13):i348–i356, 2008.

(91) CY Park, AA Klammer, L Käll, MJ MacCoss and **WS Noble**. "Rapid and accurate peptide identification from tandem mass spectra." *Journal of Proteome Research*. 7(7):3022–3027, 2008.

(92) L Käll, JD Storey and **WS Noble**. "Nonparametric estimation of posterior error probabilities associated with peptides identified by tandem mass spectrometry." *Bioinformatics. (Proceedings of the European Conference on Computational Biology)*. 24(16):i42–i48, 2008.

(93) N Yosef, R Sharan and **WS Noble**. "Improved network-based identification of protein orthologs." *Bioinformatics. (Proceedings of the European Conference on Computational Biology)*. 24(16):i200–i206, 2008. PMID18689852.

(94) S Gupta, JH Dennis, RE Thurman, RE Kingston, JA Stamatoyannopoulos and **WS Noble**. "Predicting human nucleosome occupancy from primary sequence." *PLOS Computational Biology*. 4(8):e10000134, 2008.

(95) H Lian, W Thompson, RE Thurman, JA Stamatoyannopoulos, **WS Noble** and C Lawrence. "Automated mapping of large-scale chromatin structure in ENCODE." *Bioinformatics*. 24(17):1911–1916, 2008.

(96) I Melvin, J Weston, CS Leslie and **WS Noble**. "Combining classifiers for improved classification of proteins from sequence or structure." *BMC Bioinformatics*. 9:389, 2008.

(97) GE Merrihew, CA Davis, B Ewing, G Williams, L Käll, BE Frewen, **WS Noble**, P Green, JH Thomas, MJ MacCoss. "Use of shotgun proteomics for the identification, confirmation and correction of *C. elegans* gene annotations." *Genome Research*. 18(10):1660-1669, 2008.

(98) SM Reynolds, L Käll, M Riffle, JA Bilmes and **WS Noble**. "Transmembrane topology and signal peptide prediction using dynamic Bayesian networks." *PLOS Computational Biology*. 4(11):e1000213, 2008.

(99) I Melvin, J Weston, CS Leslie and **WS Noble**. "RANKPROP: a web server for protein remote homology detection." *Bioinformatics*. 25(1):121–122, 2009.

(100) JR Hesselberth, X Chen, Z Zhang, PJ Sabo, R Sandstrom, AP Reynolds, RE Thurman, S Neph, MS Kuehn, **WS Noble**, S Fields and JA Stamatoyannopoulos. "Global mapping of protein-DNA interactions *in vivo* by digital genomic footprinting." *Nature Methods*. 6(4):283–289, 2009.

(101) L Käll, JD Storey and **WS Noble**. "QVALITY: Nonparametric estimation of q values and posterior error probabilities." *Bioinformatics*. 25(7):964–966, 2009.

(102) AA Klammer, CY Park and **WS Noble**. "Statistical calibration of the SEQUEST XCorr function." *Journal of Proteome Research*. 8(4):2106–2113, 2009.

(103) SM Reynolds, JA Bilmes and **WS Noble**. "On the relationship between DNA periodicity and local chromatin structure." *Proceedings of the International on Research in Computational Biology (RE-COMB).* May 18–21, 2009. pp. 434–450.

(104) J Hawkins, CE Grant, **WS Noble** and TL Bailey. "Assessing phylogenetic motif models for predicting transcription factor binding sites." *Bioinformatics (Proceedings of the Intelligent Systems for Molecular Biology Conference).* 25(12):i339–347, 2009.

(105) T Mann, R Humbert, MO Dorschner, JA Stamatoyannopoulos and **WS Noble**. "A thermodynamic approach to PCR primer design." *Nucleic Acids Research.* 37(13):e95, 2009.

(106) TL Bailey, M Boden, FA Buske, MC Frith, CE Grant, L Clementi, J Ren, WW Li and **WS Noble**. "MEME Suite: tools for motif discovery and searching.' *Nucleic Acids Research.* 37(Web server issue):W202-208, 2009.

(107) M Spivak, J Weston, L Bottou, L Käll and **WS Noble**. "Improvements to the Percolator algorithm for peptide identification from shotgun proteomics data sets." *Journal of Proteome Research.* 8(7):3737–3745, 2009.

(108) **WS Noble**. "A quick guide to organizing computational biology experiments." *PLOS Computational Biology.* 5(7):e1000424, 2009.

(109) **WS Noble**. "How does multiple testing correction work?" *Nature Biotechnology.* 27(12):1135-1137, 2009.

(110) SM Reynolds, Z Weng, JA Bilmes and **WS Noble**. "Predicting nucleosome positioning using multiple evidence tracks." *Proceedings of the International on Research in Computational Biology (RECOMB).* Lecture Notes in Computer Science, 2010, Volume 6044/2010, pp. 441-455.

(111) Z Aydin, JI Murray, RH Waterston and **WS Noble**. "Using machine learning to speed up manual image annotation: application to 3D imaging protocol for measuring single cell gene expression in the developing *C. elegans* embryo." *BMC Bioinformatics.* 11(1):84, 2010.

(112) M Hue, M Riffle, JP Vert and **WS Noble**. "Large-scale prediction of protein-protein interactions from structures." *BMC Bioinformatics.* 11:144, 2010.

(113) N Yosef, J Gramm, Q Wang, **WS Noble**, RM Karp and R Sharan. "Prediction of phenotype information from genotype data." *Communications in Information and Systems.* 10(2):99–114, 2010.

(114) S McIlwain, P Draghicescu, P Singh, DR Goodlett and **WS Noble**. "Detecting cross-linked peptides by searching against a database of cross-linked peptide pairs." *Journal of Proteome Research.* 9(5):2488-2495, 2010.

(115) X Chen, MM Hoffman, JA Bilmes, JR Hesselberth and **WS Noble**. "A dynamic Bayesian network for identifying protein binding footprints from single molecule based sequencing data." *Bioinformatics (Proceedings of the ISMB).* 26(12):i334–i342, 2010.

(116) MM Hoffman, OJ Buske and **WS Noble**. "The Genomedata format for storing large-scale functional genomics data." *Bioinformatics.* 26(11):1458–1459, 2010.

(117) SM Reynolds, JA Bilmes and **WS Noble**. "Learning a weighted sequence model of the nucleosome core and linker yields more accurate predictions in *Saccharomyces cerevisiae* and *Homo sapiens*." *PLOS Computational Biology.* 6(7):e10000834, 2010.

(118) E Hirsch, C Lim, D Dobrez, MG Adams and **WS Noble**. "Meteorological factors and the timing of the initiating event of parturition." *International Journal of Biometeorology.* 55(2):265, 2011.

(119) Z Duan, M Andronescu, K Schutz, S McIlwain, YJ Kim, C Lee, J Shendure, S Fields, CA Blau and **WS Noble**. "A three-dimensional model of the yeast genome." *Nature.* 465:363–367, 2010.

(120) P Agius, A Arvey, W Chang, **WS Noble** and CS Leslie. "High resolution models of transcription factor-DNA affinities improve in vitro and in vivo binding predictions." *PLOS Computational Biology.* 6(9):e10000916, 2010.

(121) O Serang, MJ MacCoss and **WS Noble**. "Efficient marginalization to compute protein posterior probabilities from shotgun mass spectrometry data." *Journal of Proteome Research.* 9(10):5346–5357, 2010.

(122) V Sharma, JK Eng, S Feldman, P von Haller, MJ MacCoss and **WS Noble**. "Precursor charge state prediction for electron transfer dissociation tandem mass spectra." *Journal of Proteome Research.* 9(10):5438–5444, 2010.

(123) I Melvin, J Weston, **WS Noble** and CS Leslie. "Detecting remote evolutionary relationships among proteins by large-scale semantic embedding." *PLOS Computational Biology.* 7(1):e1001047, 2011.

(124) CE Grant, TL Bailey and **WS Noble**. "FIMO: Scanning for occurrences of a given motif." *Bioinformatics.* 27(7):1017–1018, 2011.

(125) The ENCODE Project Consortium. "A user's guide to the Encylopedia of DNA Elements (ENCODE)." *PLOS Biology.* 9(4):e1001046, 2011.

(126) V Granholm, **WS Noble** and L Käll. "On using samples of known protein content to assess the statistical calibration of scores assigned to peptide-spectrum matches in shotgun proteomics." *Journal of Proteome Research.* 10(5):2671–2678, 2011.

(127) Z Aydin, A Singh, JA Bilmes and **WS Noble**. "Learning sparse models for a dynamic Bayesian network classifier of protein secondary structure." *BMC Bioinformatics.* 12:154, 2011.

(128) E Tanaka, TL Bailey, CE Grant, **WS Noble** and U Keich. "Improved similarity scores for comparing motifs." *Bioinformatics.* 27(12): 1603–1609, 2011.

(129) BJ Diament and **WS Noble**. "Faster SEQUEST searching for peptide identification from tandem mass spectra." *Journal of Proteome Research.* 10(9):3871–9, 2011.

(130) OJ Buske, MM Hoffman, N Ponts, KG Le Roch and **WS Noble**. "Exploratory analysis of genomic segmentations with Segtools." *BMC Bioinformatics.* 12:415, 2011.

(131) G Cuellar-Partida, FA Buske, RC McLeay, T Whitington, **WS Noble** and TL Bailey. "Epigenetic priors for identifying active transcription factor binding sites." *Bioinformatics.* 28:56-62, 2011.

(132) M Spivak, J Weston, MJ MacCoss and **WS Noble**. "Direct maximization of protein identifications from tandem mass spectra." *Molecular and Cellular Proteomics.* 11(2):M111.012161, 2012.

(133) DM Witten and **WS Noble**. "On the estimation of statistical significance of three-dimensional colocalization of sets of genomic elements." *Nucleic Acids Research.* 40(9):3849–3855, 2012.

(134) O Serang and **WS Noble**. "Faster mass spectrometry-based protein inference: junction trees are more efficient than sampling and marginalization by enumeration." *IEEE/ACM Transactions on Computational Biology and Bioinformatics.* 9(3):809–817, 2012.

(135) Y Qi, M Oja, J Weston and **WS Noble**. "A unified multitask architecture for predicting local protein properties." *PLOS One.* 7(3):e32235, 2012.

(136) MM Hoffman, OJ Buske, Z Weng, JA Bilmes and **WS Noble**. "Unsupervised pattern discovery in human chromatin structure through genomic segmentation." *Nature Methods.* 9(5):473–476, 2012.

(137) A Arvey, P Agius, **WS Noble** and CS Leslie. "Sequence and chromatin determinants of cell-type specific transcription factor binding." *Genome Research.* 22(9):1723–34, 2012.

(138) J Wang, J Zhuang, S Iyer, X Lin, TW Whitfield, M Greven, X Dong, A Kundaje, Y Cheng, E Birney, RM Myers, **WS Noble**, MP Snyder and Z Weng. "Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors." *Genome Research.* 22(9):1798–1812, 2012.

(139) The ENCODE Project Consortium. "An integrated encylopedia of DNA elements in the human genome." *Nature.* 489(7414):57–74, 2012.

(140) M Fioramonte, AM dos Santos, S McIlwain, **WS Noble**, KG Franchini and FC Gozzo. "Analysis of secondary structure in proteins by chemical cross-linking coupled to mass spectrometry." *Proteomics.* 12(17):2746–52, 2012.

(141) M Spivak, MS Bereman, MJ MacCoss and **WS Noble**. "Learning score function parameters for improved spectrum identification in tandem mass spectrometry experiments." *Journal of Proteome Research.* 11(9):4499–4508, 2012.

(142) SC Di Rienzi, KC Lindstrom, T Mann, **WS Noble**, MK Raghuraman and BJ Brewer. "Maintaining replication origins in the face of genomic change." *Genome Research.* 22(10):1940-1952, 2012.

(143) Z Aydin, J Thompson, JA Bilmes, D Baker and **WS Noble**. "Protein torsion angle class prediction by a hybrid architecture of Bayesian and neural networks." *Proceedings of the 13th International Conference on Bioinformatics and Computational Biology.* pp. 2012–2018. July 16–19, 2012.

(144) AP Singh, J Halloran, JA Bilmes and **WS Noble**. "Spectrum identification using a dynamic Bayesian network model of tandem mass spectra." *Uncertainty in Artificial Intelligence: Proceedings of the Twenty-Eighth Conference.* Aug. 15-17, 2012. pp. 775-784.

(145) S Ryu, DR Goodlett, **WS Noble** and VN Minin. "A statistical approach to peptide identification from clustered tandem mass spectrometry data." *Proceedings of the International Workshop on Computational Proteomics, IEEE BIBM.* pp. 648–653, 2012.

(146) S McIlwain, M Mathews, MS Bereman, EW Rubel, MJ MacCoss and **WS Noble**. "Estimating relative abundances of proteins from shotgun proteomics data." *BMC Bioinformatics.* 13:308, 2012.

(147) MM Hoffman, J Ernst, SP Wilder, A Kundaje, RS Harris, MW Libbrecht, B Giardine, P Ellenbogen, JA Bilmes, E Birney and RC Hardison, I Dunham, M Kellis and **WS Noble**. "Integrative annotation of chromatin elements from ENCODE data." *Nucleic Acids Research.* 41(2):827–841, 2013.

(148) MT Weirauch, A Cote, R Norel, M Annala, Y Zhao, S Talukder, A Vedenko, DREAM5 consortium, QD Morris, ML Bulyk, G Stolovitzky and TR Hughes. "An evaluation of methods for the modeling of transcription factor sequence specificity." *Nature Biotechnology.* 31(2):126–34, 2013.

(149) V Granholm, **WS Noble** and L Käll. "Determining the calibration of confidence estimation procedures for unique peptides in shotgun proteomics." *Journal of Proteomics.* 80(27):123–131, 2012.

(150) X Deng, JB Berletch, W Ma, DK Nguyen, JB Hiatt, **WS Noble**, J Shendure and CM Disteche. "Mammalian X upregulation is associated with enhanced transcription initiation, RNA half-life, and MOF-mediated H4K16 acetylation." *Developmental Cell.* 25(1):55–68, 2013.

(151) DA Skelly, GE Merrihew, M Riffle, CF Connelly, EO Kerr, M Johansson, D Jaschob, B Graczyk, NJ Shulman, J Wakefield, SJ Cooper, S Fields, **WS Noble**, EGD Muller, TN Davis, MJ Dunham, MJ MacCoss and JM Akey. "Integrative phenomics reveals insight into the structure of phenotypic diversity in budding yeast." *Genome Research.* 23(9):1496–1504, 2013.

(152) MJ Zeitz, F Ay, JD Heidmann, PL Lerner, **WS Noble**, BN Steelman and AR Hoffman. "Genomic interaction profiles in breast cancer reveal altered chromatin architecture." *PLOS One.* 8(9):e73974, 2013.

(153) MJ Zeitz, PL Lerner, <u>F Ay</u>, E Van Nostrand, JD Heidmann, **WS Noble** and AR Hoffman. "Implications of COMT long-range interactions on the phenotypic variability of 22q11.2 deletion syndrome." *Nucleus.* 4(6):487–493, 2013.

(154) BS Sexton, D Avey, BR Druliner, JA Fincher, DL Vera, DJ Grau, ML Borowsky, <u>S Gupta</u>, S Girimurugan, E Chicken, J Zhang, **WS Noble**, F Zhu, RE Kingston, and JH Dennis. "The spring-loaded genome: Nucleosome redistributions are widespread, transient, and DNA-directed." *Genome Research.* 24(2):251-259, 2013.

(155) <u>F Ay</u>, EM Bunnik, N Varoquaux, SM Bol, J Prudhomme, JP Vert, **WS Noble** and KG Le Roch. "Three-dimensional modeling of the *P. falciparum* genome during the erythrocytic cycle reveals a strong connection between genome architecture and gene expression." *Genome Research.* 24(6):974–988, 2014.

(156) <u>F Ay</u>, TL Bailey and **WS Noble**. "Statistical confidence estimation for Hi-C data reveals regulatory chromatin contacts." *Genome Research.* 24(6):999–1011, 2014.

(157) <u>W Ma</u>, **WS Noble** and TL Bailey. "Motif-based analysis of large nucleotide datasets using MEME-ChIP." *Nature Protocols.* 9(6):1428–1450, 2014.

(158) N Varoquaux, <u>F Ay</u>, **WS Noble** and JP Vert. "A statistical approach for inferring the 3D structure of the genome." *Bioinformatics (Proceedings of the ISMB).* 30(12):i26–i33, 2014.

(159) PJ Park, J Ho, Y Jung, T Liu, B Alver, S Lee, K Ikegami, K Sohn, A Minoda, M Tolstorukov, A Appert, S Parker, T Gu, A Kundaje, N Riddle, E Bishop, T Egelhofer, S Hu, A Alekseyenko, Y Schwartz, D Asker, J Belsky, S Bowman, Q Chen, R Chen, D Day, Y Dong, X Duan, CB Epstein, S Ercan, E Feingold, F Ferrari, J Garrigues, N Gehlenborg, PJ Good, P Haseley, D He, M Herrmann, <u>MM Hoffman</u>, T Jeffers, P Kharchenko, P Kolasinska-Zwierz, C Kotwaliwale, N Kumar, S Langley, E Larschan, I Latorre, <u>MW Libbrecht</u>, X Lin, R Park, M Pazin, H Pham, A Plachetka, B Qin, N Shoresh, P Stempor, A Vielle, C Wang, C Whittle, H Xue, R Kingston, JH Kim, BE Bernstein, A Dernburg, V Pirrotta, M Kuroda, **WS Noble**, T Tullius, M Kellis, D MacAlpine, S Strome, S Elgin, XS Liu, J Lieb, J Ahringer, G Karpen, A Rechtsteiner and A Dose. "Comparative analysis of metazoan chromatin organization." *Nature.* 512:449-452, 2014.

(160) <u>H Zare</u>, J Wang, <u>A Hu</u>, <u>K Weber</u>, J Smith, D Nickerson, C Song, D Witten, CA Blau and **WS Noble**. "Inferring clonal composition from multiple sections of a breast cancer." *PLOS Computational Biology.* 10(7):e1003703, 2014.

(161) <u>JT Halloran</u>, JA Bilmes and **WS Noble**. "Learning peptide-spectrum alignment models for tandem mass spectrometry." Thirtieth Conference on Uncertainty in Artificial Intelligence (UAI 2014). Ed. Nevin Zhang and Jin Tian. AUAI Press, Corvallis, OR. pp. 320-329, 2014.

(162) <u>JJ Howbert</u> and **WS Noble**. "Computing exact p-values for a cross-correlation shotgun proteomics scoring function." *Molecular and Cellular Proteomics.* 13(9):2467–2479, 2014.

(163) <u>S McIlwain</u>, <u>K Tamura</u>, <u>A Kertesz-Farkas</u>, <u>CE Grant</u>, <u>BJ Diament</u>, <u>BE Frewen</u>, <u>JJ Howbert</u>, MR Hoopmann, L Käll, JK Eng, MJ MacCoss, **WS Noble**. "Crux: rapid open source protein tandem mass spectrometry analysis." *Journal of Proteome Research.* 13(10):4488–4491, 2014.

(164) <u>W Ma</u>, F Ay, C Lee, G Gulsoy, X Deng, S Cook, J Hesson, C Cavanaugh, CB Ware, A Krumm, J Shendure, CA Blau, CM Disteche, **WS Noble**, Z Duan. "Fine-scale chromatin interaction maps reveal the cis-regulatory landscape of lincRNA genes in human cells." *Nature Methods.* 12(1):1–91, 2015.

(165) <u>F Ay</u>, EM Bunnik, N Varoquaux, JP Vert, **WS Noble** and KG Le Roch. "Multiple dimensions of epigenetic gene regulation in the malaria parasite *Plasmodium falciparum*." *Bioessays.* 37(2):182–194, 2015.

(166) U Keich and **WS Noble**. "On the importance of well calibrated scores for identifying shotgun proteomics spectra." *Journal of Proteome Research.* 14(2):1147–1160, 2015.

(167) F Ay, TH Vu, MJ Zeitz, N Varoquaux, JE Carette, JP Vert, AR Hoffman and **WS Noble**. "Identifying multi-locus chromatin contacts in human cells using tethered multiple 3C." *BMC Genomics.* 16(1):121, 2015.

(168) F Yang, X Deng, W Ma, JB Berletch, N Rabaia, JM Moore, GN Filippova, Y Liu, **WS Noble**, J Shendure and CM Disteche. "The lncRNA Firre anchors the inactive X chromosome to the nucleolus by binding CTCF and maintains H3K27me3 methylation." *Genome Biology.* 16(1):52, 2015.

(169) Z Aydin, D Baker and **WS Noble**. "Constructing structural profiles for protein torsion angle prediction." *Proceedings of the International Conference on Bioinformatics Models, Methods and Algorithms (BIOINFORMATICS-2015).* Lisbon, Portugal, pp. 26–35, 2015.

(170) MW Libbrecht, F Ay, MM Hoffman, DM Gilbert, JA Bilmes and **WS Noble**. "Joint annotation of chromatin state and chromatin conformation reveals relationships among domain types and identifies domains of cell type-specific expression." *Genome Research.* 25(4):544–57, 2015. *Selected as one of the top 10 regulatory genomics published papers in 2014–2015 by the ISCB Special Interest Group On Regulatory & Systems Genomics*

(171) JB Berletch, W Ma, F Yang, J Shendure, **WS Noble**, CM Disteche, X Deng. "Escape from X inactivation varies in mouse tissues." *PLOS Genetics.* 18(3):e1005079, 2015.

(172) MW Libbrecht and **WS Noble**. "Machine learning applications in genetics and genomics." *Nature Reviews Genetics.* 16(6):321-332, 2015.

(173) N Varoquaux, I Liachko, F Ay, J Burton, J Shendure, MJ Dunham, JP Vert and **WS Noble**. "Accurate identification of centromere locations in yeast genomes using Hi-C." *Nucleic Acids Research.* 43(11):5331–5339, 2015.

(174) TL Bailey, J Johnson, CE Grant and **WS Noble**. "The MEME Suite." *Nucleic Acids Research.* 43(W1):W39–49, 2015.

(175) MW Libbrecht, MM Hoffman, JA Bilmes, **WS Noble**. "Entropic graph-based posterior regularization." *Proceedings of the 32nd International Conference on Machine Learning.* Lille, France. pp. 1992-2001, 2015.

(176) JK Eng, MR Hoopmann, TA Jahan, JD Egertson, **WS Noble**, and MJ MacCoss. "A deeper look into Comet—implementation and features." *Journal of Proteome Research.* 26(11):1865–1874, 2015.

(177) M Riffle, GE Merrihew, D Jaschob, V Sharma, TN Davis, **WS Noble**, MJ MacCoss. "Visualization and dissemination of multidimensional proteomics data comparing protein abundance during *Caenorhabditis elegans* development." *Journal of the American Society for Mass Spectrometry.* 26(11):1827–1836, 2015.

(178) **WS Noble**. "Mass spectrometrists should only search for peptides they care about." *Nature Methods.* 12(7):605–608, 2015.

(179) V Dileep, F Ay, J Sima, DL Vera, **WS Noble**, and DM Gilbert. "Topologically-associating domains and their long-range contacts are established during early G1 coincident with the establishment of the replication timing program." *Genome Research.* 25(8):1104–1113, 2015.

(180) A Kertesz-Farkas, U Keich and **WS Noble**. "Tandem mass spectrum identification via cascade search." *Journal of Proteome Research.* 14(8):3027–3038, 2015.

(181) RM Gittelman, E Hun, F Ay, J Madeoy, L Pennacchio, **WS Noble**, DR Hawkins and JM Akey. "Comprehensive identification and analysis of human accelerated regulatory DNA." *Genome Research.* 25(9):1245–1255, 2015.

(182) U Keich, A Kertesz-Farkas and **WS Noble**. "Improved false discovery rate estimation procedure for shotgun proteomics." *Journal of Proteome Research.* 14(8):3148–3161, 2015.

(183) F Ay and **WS Noble**. "Analysis methods for 3D nucleome data." *Genome Biology.* 16(1):183, 2015.

(184) X Deng, W Ma, V Ramani, AJ Hill, F Yang, F Ay, JB Berletch, CA Blau, J Shendure, Z Duan, **WS Noble**, CM Disteche. "Bipartite structure of the inactive mouse X chromosome." *Genome Biology.* 16(1):152, 2015.

(185) YS Ting, JD Egertson, SH Payne, S Kim, B MacLean, L Käll, R Aebersold, RD Smith, **WS Noble** and MJ MacCoss. "Peptide-centric proteome analysis: an alternative strategy for the analysis of tandem mass spectrometry data." *Molecular and Cellular Proteomics.* 14(9):2301-2307, 2015.

(186) CE Grant, J Johnson, TL Bailey, and **WS Noble**. "MCAST: Scanning for *cis*-regulatory motif clusters." *Bioinformatics.* 32(8):1217–1219, 2015.

(187) JB Berletch, W Ma, F Yang, J Shendure, **WS Noble**, and CM Disteche. "Identification of genes escaping X inactivation by allelic expression analysis in a novel hybrid mouse model." *Data in Brief.* 5:761-769, 2015.

(188) S Wang, JT Halloran, JA Bilmes and **WS Noble**. "Faster and more accurate graphical model identification of tandem mass spectra using trellises." *Bioinformatics (Proceedings of the ISMB).* 32(12):i322–i331, 2016.

(189) OK Smith, R Kim, H Fu, MM Martin, K Utani, Y Zhang, AB Marks, M Lalande, S Chamberlain, MW Libbrecht, EE Bouhassira, MC Ryan, **WS Noble** and MI Aladjem. "Distinct epigenetic features of differentiation-regulated replication origins." *Epigenetics and Chromatin.* 9(1):18, 2016.

(190) DH May, E Timmins-Schiffman, MP Mikan, HR Harvey, E Borenstein, BL Nunn and **WS Noble**. "An alignment-free 'metapeptide' strategy for metaproteomic characterization of microbiome samples using shotgun metagenomic sequencing." *Journal of Proteome Research.* 15(8):2697–2705, 2016.

(191) J Halloran, JA Bilmes and **WS Noble**. "A dynamic Bayesian network for accurate detection of peptides from tandem mass spectra." *Journal of Proteome Research.* 15(8):2749–2759, 2016.

(192) M The, MJ MacCoss, **WS Noble** and L Käll. "Fast and accurate protein false discovery rates on large-scale proteomics data sets with Percolator 3.0." *Journal of the American Society of Mass Spectrometry.* 27(11):1719–1727, 2016.

(193) V Ramani, DA Cusanovich, RJ Hause, W Ma, X Deng, CA Blau, CM Disteche, **WS Noble**, J Shendure, Z Duan. "Mapping three-dimensional genome architecture through in situ DNase Hi-C." *Nature Protocols.* 11(11):2014–2121, 2016.

(194) E Timmins-Schiffman, DH May, M Mikan, M Riffle, C Frazar, HR Harvey, **WS Noble**, BL Nunn. "Critical decisions in metaproteomics: Achieving high confidence protein annotations in a sea of unknowns." *The ISME Journal.* 11(2):309–314, 2017.

(195) W Bai, JA Bilmes and **WS Noble**. "Bipartite matching generalizations for peptide identification in tandem mass spectrometry." *ACM Conference on Bioinformatics, Computational Biology, and Health Informatics.* Seattle, WA. 327–336, 2016. Winner of Best Paper award.

(196) K Wei, MW Libbrecht, JA Bilmes and **WS Noble**. "Choosing panels of genomics assays using submodular optimization." *Genome Biology.* 17(1):229, 2016.

(197) V Ramani, X Deng, R Qiu, KL Gunderson, FJ Steemers, CM Disteche, **WS Noble**, Z Duan, J Shendure. "Massively multiplex single-cell Hi-C." *Nature Methods.* 14(3):263–266, 2017.

(198) U Keich and **WS Noble**. "Progressive calibration and averaging for tandem mass spectrometry statistical confidence estimation: Why settle for a single decoy?" *Proceedings of the International Conference on Research in Computational Biology (RECOMB).* 10229:99-116, 2017.

(199) MN Gabere and **WS Noble**. "Empirical comparison of web-based antimicrobial peptide prediction tools." *Bioinformatics.* 33(13):1921–1929, 2017.

(200) S Wang, H Cai, JA Bilmes, and **WS Noble**. "Training compressed fully-connected networks with a density-diversity penalty." *Proceedings of the Fifth International Conference on Learning Representations.* 2017.

(201) U Eser, D Chandler-Brown, F Ay, AF Straight, Z Duan, **WS Noble**, JM Skotheim. "Form and function of topologically associating genomic domains in budding yeast." *Proceedings of the National Academy of Science.* 114(15):E3061–E3070, 2017.

(202) GG Yardımcı and **WS Noble**. "Software tools for visualizing Hi-C data." *Genome Biology.* 18(1):26, 2017.

(203) ZE Sychev, A Hu, T DiMaio, A Gitter, ND Camp, **WS Noble**, A Wolf-Yadlin, M Lagunoff. "Integrated systems biology analysis of KSHV latent infection reveals viral induction and reliance on peroxisome mediated lipid metabolism." *PLOS Pathogens.* 13(3):e1006256, 2017.

(204) DH May, K Tamura, **WS Noble**. "Param-Medic: A tool for improving MS/MS database search yield by optimizing parameter settings." *Journal of Proteome Research.* 16(4):1817–1824, 2017.

(205) K-K Yan, GG Yardımcı, C Yan, **WS Noble**, M Gerstein. "HiC-Spector: a matrix library for spectral and reproducibility analysis of Hi-C contact maps." *Bioinformatics.* 33(14):2199–2201, 2017.

(206) W Ma, L Yang, R Rohs and **WS Noble**. "DNA sequence+shape kernel enables alignment-free modeling of transcription factor binding." *Bioinformatics.* 33(19):3003–3010, 2017.

(207) S Kim, I Liachko, DG Brickner, KB Cook, **WS Noble**, JH Brickner, J Shendure, MJ Dunham. "The dynamic three-dimensional organization of the diploid yeast genome." *eLife.* 6:e23623, 2017.

(208) **WS Noble** and U Keich. "Response to 'Mass spectrometrists should search for all peptides, but assess only the ones they care about.'" *Nature Methods.* 14(7):644, 2017.

(209) H Sakano, DAR Zorio, X Wang, YS Ting, **WS Noble**, MJ MacCoss, EW Rubel, Y Wang. "Proteomic analyses of nucleus laminaris identified candidate targets of the fragile X mental retardation protein." *Journal of Comparative Neurology.* 525(15):3341-3359, 2017.

(210) JM Schreiber and **WS Noble**. "Finding the optimal Bayesian network given a constraint graph." *PeerJ Computer Science.* 3:e122, 2017.

(211) YS Ting, JD Egertson, JG Bollinger, BC Searle, SH Payne, **WS Noble**, and MJ MacCoss. "PECAN: library free peptide detection for data-independent acquisition tandem mass spectrometry data." *Nature Methods.* 14(9):903–908, 2017.

(212) T Yang, F Zhang, GG Yardımcı, F Song, RC Hardison, **WS Noble**, F Yue, and Q Li. "HiCRep: assessing the reproducibility of Hi-C data using a stratum-adjusted correlation coefficient." *Genome Research.* 27(11):1939–1949, 2017.

(213) U Keich and **WS Noble**. "Controlling the FDR in imperfect matches to an incomplete database." *Journal of the American Statistical Association.* 113(523):973–982, 2018.

(214) **WS Noble**. "Ten simple rules for writing a response to reviewers." *PLOS Computational Biology.* 13(10):e1005730, 2017.

(215) J Dekker, AS Belmont, M Guttman, VO Leshyk, JT Lis, S Lomvardas, LA Mirny, CC O'Shea, PJ Park, B Ren, JCR Politz, J Shendure, S Zhong, and the 4D Nucleome Network. "The 4D Nucleome Project." *Nature* 549(7671):219–226, 2017.

(216) RCW Chan, MW Libbrecht, EG Roberts, JA Bilmes, **WS Noble**, and MM Hoffman. "Segway 2.0: Gaussian mixture models and minibatch training." *Bioinformatics.* 34(4):669-671, 2018.

(217) J Liu, JT Halloran, JA Bilmes, RM Daza, C Lee, EM Mahen, D Prunkard, C Song, S Blau, MO Dorschner, VK Gadi, J Shendure, CA Blau, and **WS Noble**. "Comprehensive statistical inference of the clonal structure of cancer from multiple biopsies." *Scientific Reports.* 7(1):16943, 2017.

(218) M Riffle, DH May, E Timmins-Schiffman, MP Mikan, D Jaschob, **WS Noble**, and BL Nunn. "MetaGOmics: A web-based tool for peptide-centric functional and taxonomic analysis of metaproteomics data." *Proteomes.* 6(1), 2017.

(219) MW Libbrecht, JA Bilmes and **WS Noble**. "Choosing non-redundant representative subsets of protein sequence data sets using submodular optimization." *Proteins.* 86(4):454–466, 2018.

(220) W Ma, F Ay, C Lee, G Gulsoy, X Deng, S Cook, J Hesson, C Cavanaugh, CB Ware, A Krumm, J Shendure, CA Blau, CM Disteche, **WS Noble**, Z Duan. "Using DNase Hi-C techniques to map global and local three-dimensional genome architecture at high resolution." *Methods.* 142:59–73, 2018.

(221) TJ Durham, MW Libbrecht, JJ Howbert, JA Bilmes and **WS Noble**. "PREDICTD: PaRallel Epigenomics Data Imputation with Cloud-based Tensor Decomposition." *Nature Communications.* 9(1):1402, 2018.

(222) G Bonora, X Deng, H Fang, V Ramani, R Qiu, JB Berletch, GN Filippova, Z Duan, J Shendure, **WS Noble**, CM Disteche. "Orientation-dependent Dxz4 contacts shape the 3D structure of the inactive X chromosome." *Nature Communications.* 9(1):1445, 2018.

(223) J Liu, D Lin, GG Yardımcı, and **WS Noble**. "Unsupervised embedding of single-cell Hi-C data." *Bioinformatics (Proceedings of the ISMB).* 34(13):i96–i104, 2018.

(224) O Ursu, N Boley, M Taranova, YXR Wang, GG Yardımcı, **WS Noble**, and A Kundaje. "GenomeDISCO: A concordance score for chromosome conformation capture experiments using random walks on contact map graphs." *Bioinformatics.* 34(16):2701–2707, 2018.

(225) EM Bunnik, KB Cook, N Varoquaux, G Batugedara, J Prudhomme, A Cort, L Shi, C Andolina, LS Ross, D Brady, DA Fidock, F Nosten, R Tewari, P Sinnis, F Ay, JP Vert, **WS Noble**, KG Le Roch. "Changes in genome organization of parasite-specific gene families during the *Plasmodium* transmission stages." *Nature Communications.* 9(1):1910, 2018.

(226) J Dixon, J Xu, V Dileep, Y Zhan, F Song, VT Le, GG Yardımcı, A Chakraborty, DV Bann, Y Wang, R Clark, L Zhang, H Yang, T Liu, S Iyyanki, L An, C Pool, T Sasaki, JC Rivera-Mulia, H Ozadam, BR Lajoie, R Kaul, M Buckley, K Lee, M Diegel, D Pezic, C Ernst, S Hadjur, DT Odom, JA Stamatoyannopoulos, JR Broach, RC Hardison, F Ay, **WS Noble**, J Dekker, DM Gilbert, and F Yue. "Integrative detection and analysis of structural variation in cancer genomes." *Nature Genetics.* 50(10):1388–1398, 2018.

(227) Y Lu, Y Fan, J Lv, and **WS Noble**. "DeepPINK: reproducible feature selection in deep neural networks." *Advances in Neural Information Processing Systems.* 31:8690–8700, 2018.

(228) W Bai, **WS Noble**, JA Bilmes. "Submodular maximization via gradient ascent: The case of deep submodular functions." *Advances in Neural Information Processing Systems.* 31:7989–7999, 2018.

(229) W Bittremieux, P Meysman, **WS Noble**, K Laukens. "Fast open modification spectral library searching through approximate nearest neighbor indexing." *Journal of Proteome Research.* 17(10):3463–3474, 2018.

(230) A Lin, JJ Howbert, and **WS Noble**. "Combining high resolution and exact calibration to boost statistical power: A well-calibrated score function for high-resolution MS2 data." *Journal of Proteome Research.* 2018. 17(11):3644–3656, 2018.

(231) W Ma, G Bonora, JB Berletch, X Deng, **WS Noble**, CM Disteche. "X-chromosome inactivation and escape from X inactivation in mouse." *Methods in Molecular Biology.* 1861:205–219, 2018.

(232) L Pino, BC Searle, E Huang, **WS Noble**, AN Hoofnagle, MJ MacCoss. "Calibration using a single-point external reference material harmonizes quantitative mass spectrometry proteomics data between platforms and laboratories." *Analytical Chemistry.* 90(21):13112–13117, 2018.

(233) D Lin, G Bonora, GG Yardımcı, and **WS Noble**. "Computational methods for analyzing and modeling genome structure and organization." *WIREs Systems Biology*. 11(1):e1435, 2019.

(234) A Hu, Y Lu, JA Bilmes, **WS Noble**. "Joint precursor elution profile inference via regression for peptide detection in data-independent acquisition mass spectra." *Journal of Proteome Research*. 18(1):86–94, 2019.

(235) W Bai, JA Bilmes and **WS Noble**. "Submodular generalized matching for peptide identification in tandem mass spectrometry." *IEEE Transactions in Computational Biology and Bioinformatics*. 16(4):1168–1181, 2019.

(236) GG Yardımcı, H Ozadam, MEG Sauria, O Ursu, K Yan, T Yang, A Chakraborty, A Kaul, BR Lajoie, F Song, Y Zhan, F Ay, M Gerstein, A Kundaje, Q Li, J Taylor, F Yue, J Dekker, **WS Noble**. "Measuring the reproducibility and quality of Hi-C data." *Genome Biology*. 20(1):57, 2019.

(237) A Cheng, CE Grant, **WS Noble**, and TL Bailey. "MoMo: Discovery of statistically significant post-translational modification motifs." *Bioinformatics*. 35(16):2774–2782, 2019.

(238) U Keich, K Tamura, and **WS Noble**. "Averaging strategy to reduce variability in target-decoy estimates of false discovery rate." *Journal of Proteome Research*. 18(2):585–593, 2019.

(239) DH May, K Tamura, **WS Noble**. "Detecting modifications in proteomics experiments with Param-Medic." *Journal of Proteome Research*. 18(4):1902–1906, 2019.

(240) A Bertero, PA Fields, V Ramani, G Bonora, GG Yardımcı, H Reinecke, L Pabon, **WS Noble**, J Shendure, CE Murry. "Dynamics of genome reorganization during human cardiogenesis reveal an RBM20-dependent splicing factory." *Nature Communications*. 10(1):1538, 2019.

(241) M Gasperini, AJ Hill, JL McFaline-Figueroa, B Martin, S Kim, MD Zhang, D Jackson, A Leith, JM Schreiber, **WS Noble**, C Trapnell, N Ahituv, J Shendure. "A genome-wide framework for mapping gene regulation via cellular genetic screens." *Cell*. 176(1-2):377–390, 2019.

(242) JS Kim, X He, J Liu, Z Duan, T Kim, J Gerard, B Kim, MM Pillai, WS Lane, **WS Noble**, B Budnik and T Waldman. "Systematic proteomics of endogenous human cohesin reveals an interaction with diverse splicing factors and RNA binding proteins required for mitotic progression." *Journal of Biological Chemistry*. 294(22):8760–8772, 2019.

(243) W Chen, A McKenna, JM Schreiber, M Haeussler, Y Yin, V Agrawal, **WS Noble**, J Shendure. "Massively parallel profiling and predictive modeling of the outcomes of CRISPR/Cas9-mediated double-strand break repair." *Nucleic Acids Research*. 47(15):7989–8003, 2019.

(244) J Liu, Y Huang, R Singh, JP Vert, **WS Noble**. "Jointly embedding multiple single-cell omics measurements." *19th International Workshop on Algorithms in Bioinformatics (WABI 2019)*, 143(10):1–13, 2019.

(245) G Cauer, GG Yardımcı, JP Vert, N Varoquaux, **WS Noble**. "Inferring diploid 3D chromatin structures from Hi-C data." *19th International Workshop on Algorithms in Bioinformatics (WABI 2019)*. 143(11):1–13, 2019.

(246) A Bertero, PA Fields, AST Smith, A Leonard, K Beussman, NJ Sniadecki, D-H Kim, HF Tse, L Pabon, J Shendure, **WS Noble**, CE Murry. "Chromatin compartment dynamics in a haploinsufficient model of cardiac laminopathy." *Journal of Cell Biology*. 218(9):2919–2944, 2019.

(247) JT Halloran, H Zhang, K Kara, C Renggli, M The, C Zhang, DM Rocke, L Käll, **WS Noble**. "Speeding up Percolator." *Journal of Proteome Research*. 18(9):3353–3359, 2019.

(248) MW Libbrecht, O Rodriguez, Z Weng, JA Bilmes, MM Hoffman, **WS Noble**. "A unified encyclopedia of human functional DNA elements through fully automated annotation of 164 human cell types." *Genome Biology*. 20:180, 2019

(249) M Mikan, HR Harvey, E Timmins-Schiffman, M Riffle, <u>DH May</u>, I Salter, **WS Noble**, BL Nunn. "Metaproteomics reveal that rapid perturbations in organic matter inputs stimulate functional restructuring in Western Arctic Ocean microbiomes." *The ISME Journal.* 14:39–52, 2020.

(250) <u>DF Read</u>, <u>KB Cook</u>, <u>YY Lu</u>, KG Le Roch, and **WS Noble**. "Predicting gene expression in the human malaria parasite *Plasmodium falciparum.*" *PLOS Computational Biology.* 15(9):e1007329, 2019.

(251) <u>W Bittremieux</u>, K Laukens, **WS Noble**. "Extremely fast and accurate open modification spectral library searching of high-resolution mass spectra using feature hashing and graphics processing units." *Journal of Proteome Research.* 18(10):3972–3979, 2019.

(252) V Ramani, X Deng, R Qiu, CM Disteche, **WS Noble**, J Shendure, Z Duan. "Sci-Hi-C: a single-cell Hi-C method for mapping 3D genome organization in large numbers of single cells." *Methods.* 170:61–68, 2020.

(253) <u>KB Cook</u>, KG Le Roch, JP Vert, **WS Noble**. "Measuring significant changes in chromatin conformation with ACCOST." *Nucleic Acids Rsearch.* 48(5):2303–2311, 2020.

(254) Y Yin, Y Jian, KG Lam, JB Berletch, CM Disteche, **WS Noble**, FJ Steemers, RD Camerini-Otero, AC Adey, J Shendure. "High-throughput single cell genome sequencing reveals cell-autonomous equational segregation in meiosis I." *Molecular Cell.* 76(4):676–690, 2019.

(255) K Emery, S Hasam, U Keich and **WS Noble**. "Multiple competition based FDR control and peptide detection." *Proceedings of the International Conference on Research in Computational Biology (RECOMB).* 24:54–71, 2020.

(256) <u>W Fondrie</u>, **WS Noble**. "A machine learning strategy that leverages large datasets to boost statistical power in small-scale experiments." *Journal of Proteome Research.* 19(3):1147–1153, 2020.

(257) <u>L Pino</u>, BC Searle, H-Y Yang, AN Hoofnagle, **WS Noble**, MJ MacCoss. "Matrix-matched calibration curves for assessing analytical figures of merit in quantitative proteomics." *Journal of Proteome Research.* 19(3):1147–1153, 2020.

(258) S Pan, MAJ Hullar, LA Lai, H Peng, <u>DH May</u>, **WS Noble**, D Raftery, SL Navarro, ML Neuhouser, PD Lampe, JW Lampe, R Chen. "Gut microbial protein expression in response to dietary patterns in a controlled feeding study: a metaproteomic approach." *Microorganisms.* 8(3):379, 2020.

(259) <u>JM Schreiber</u>, <u>TJ Durham</u>, JA Bilmes, **WS Noble**. "Multi-scale deep tensor factorization learns a latent representation of the human epigenome." *Genome Biology.* 21:81, 2020.

(260) <u>JM Schreiber</u>, JA Bilmes, **WS Noble**. "Completing the ENCODE3 compendium yields accurate imputations across a variety of assays and human biosamples." *Genome Biology.* 21:82, 2020.

(261) A Erijman, L Kozlowski, S Sohrabi-Jahromi, J Fishburn, L Warfield, <u>JM Schreiber</u>, **WS Noble**, J Söding, S Hahn. "A high-throughput screen for transcription activation domains reveals their sequence characteristics and permits reliable prediction by deep learning." *Molecular Cell.* 78(5):P890–P892, 2020.

(262) ENCODE Consortium. "Expanded encyclopedias of DNA elements in the human and mouse genomes." *Nature.* 583(7818):699–710, 2020.

(263) <u>JM Schreiber</u>, JA Bilmes, **WS Noble**. "apricot: Submodular selection for data summarization in Python." *Journal of Machine Learning Research.* 21(161):1–6, 2020

(264) <u>JM Schreiber</u>, D Hegde, **WS Noble**. "Zero-shot imputations across species are enabled through joint modeling of human and mouse epigenomics." *BCB '20: Proceedings of the 11th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics.* Article 39, pages 1–9, 2020. Winner of Best Paper award.

(265) R Singh, P Demetci, <u>G Bonora</u>, V Ramani, C Lee, H Fang, Z Duan, X Deng, J Shendure, CM Disteche, **WS Noble**. "Unsupervised manifold alignment for single-cell multi-omics data." *BCB '20: Proceedings of the 11th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics.* Article 40, pages 1–10, 2020.

(266) W Yang, JA Bilmes, **WS Noble**. "Submodular sketches of single-cell RNA-seq measurements" *ACM-BCB '20: Proceedings of the 11th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics.* Article 61, pages 1–6, 2020.

(267) <u>H-J Kim</u>, <u>GG Yardımcı</u>, <u>G Bonora</u>, V Ramani, <u>J Liu</u>, R Qiu, C Lee, J Hesson, CB Ware, J Shendure, Z Duan and **WS Noble**. "Capturing cell type-specific chromatin structural patterns by applying topic modeling to single-cell Hi-C data." *PLOS Computational Biology.* 16(9):e1008173, 2020.

(268) H Fang,[†] <u>G Bonora</u>,[†] JP Lewandowski, J Thakur, GN Filippova, S Henikoff, J Shendure, Z Duan, JL Rinn, X Deng, **WS Noble**, CM Disteche. "Trans- and cis-acting effects of the lncRNA Firre on epigenetic and structural features of the inactive X chromosome." *Nature Communications.* 11(1):6053, 2020

(269) <u>JM Schreiber</u>, <u>R Singh</u>, JA Bilmes and **WS Noble**. "A pitfall for machine learning methods aiming to predict across cell types." *Genome Biology.* 21(1):282, 2020.

(270) <u>JM Schreiber</u>, JA Bilmes, **WS Noble**. "Prioritizing transcriptomic and epigenomic experiments by using an optimization strategy that leverages imputed data." *Bioinformatics.* 37(4):439-447, 2021.

(271) P Demetci, R Santorella, B Sandstede, **WS Noble**, R Singh. "Gromov-Wasserstein optimal transport to align single-cell multi-omics data." *Proceedings of the International Conference on Research in Computational Biology (RECOMB).* 2021.

(272) <u>D Lin</u>, <u>J Sanders</u>, **WS Noble**. "HiCRep.py: Fast comparison of Hi-C contact matrices in Python." *Bioinformatics.* 37(18):2996–2997, 2021.

(273) <u>WE Fondrie</u>, **WS Noble**. "mokapot: Fast and flexible semi-supervised learning for peptide detection." *Journal of Proteome Research.* 20(4):1966–1971, 2021.

(274) <u>YY Lu</u>, JA Bilmes, RA Rodriguez-Mias, J Villén, **WS Noble**. "DIAmeter: Matching peptides to data-independent acquisition mass spectrometry data." *Bioinformatics (Proceedings of the ISMB).* 37(Supplement_1):i434–i442, 2021.

(275) M Mudge, BL Nunn, E Firth, M Ewert, <u>K Hales</u>, <u>WE Fondrie</u>, **WS Noble**, J Toner, B Light, KA Junge. "Subzero, saline incubations of Colwellia psychrerythraea reveal strategies and biomarkers for sustained life in extreme icy environments." *Environmental Microbiology.* 23(7):3840–3866, 2021.

(276) <u>TJ Durham</u>, RM Daza, L Gevirtzman, DA Cusanovich, **WS Noble**, J Shendure, RH Waterston. "Comprehensive characterization of tissue-specific chromatin accessibility in L2 *Caenorhabditis elegans* nematodes." *Genome Research.* 31:1952–1969, 2021.

(277) <u>YY Lu</u>, W Guo, X Xing, **WS Noble**. "Robust decoy-enhanced saliency maps." *Proceedings of the International Conference on Machine Learning (ICML).* 2021.

(278) <u>YY Lu</u>, <u>TC Yu</u>, <u>G Bonora</u>, **WS Noble**. "ACE: Explaining cluster from an adversarial perspective." *Proceedings of the International Conference on Machine Learning (ICML).* 139:7156–7167, 2021.

(279) <u>G Bonora</u>, V Ramani, <u>R Singh</u>, H Fang, D Jackson, S Srivatsan, R Qiu, C Lee, C Trapnell, J Shendure, Z Duan, X Deng, **WS Noble**, CM Disteche. "Single-cell landscape of nuclear configuration and gene expression during stem cell differentiation and X inactivation." *Genome Biology.* 22:279, 2021.

(280) W Bittremieux, K Laukens, **WS Noble**, PC Dorrestein. "Large-scale tandem mass spectrum clustering using fast nearest neighbor searching." *Rapid Communications in Mass Spectrometry.* e9153, 2021.

[†]Co-first authors

(281) <u>A Lin</u>, D Plubell, U Keich, **WS Noble**. "Accurately assigning peptides to spectra when only a subset of peptides are relevant." *Journal of Proteome Research.* 20(8):4153–4164, 2021.

(282) <u>W Fondrie</u>, W Bittremieux, **WS Noble**. "ppx: Programmatic access to proteomics data repositories." *Journal of Proteome Research.* 20(9):4621–4624, 2021.

(283) SG Swygert, <u>D Lin</u>, S Portillo-Ledesma, P-Y Lin, DR Hunt, C-F Kao, T Schlick, **WS Noble**, T Tsukiyama. "Local chromatin fiber folding represses transcription and loop extrusion in quiescent cells." *eLife.* e72062, 2021.

(284) <u>YY Lu</u>, **WS Noble**. "A wider field of view to predict expression." *Nature Methods.* 18:1155–1160, 2021

(285) P Demetci, R Santorella, B Sandstede, **WS Noble**, R Singh. "SCOT: single-cell multi-omics alignment with optimal transport." *Journal of Computational Biology.* 29(1):3–18, 2022.

(286) P Demetci, R Santorella, B Sandstede, **WS Noble**, R Singh. "Single-Cell Multiomics Integration by SCOT." *Journal of Computational Biology.* 29(1):19–22, 2022.

(287) WS Phipps, K Smith, H-Y Yang, CM Henderson, H Pflaum, M Lerch, <u>WE Fondrie</u>, M Emrick, MJ MacCoss, **WS Noble**, AN Hoofnagle. "Tandem mass spectrometry-based amyloid typing using manual microdissection and open-source data processing." *American Journal of Clinical Pathology.* 157(5):748–757, 2022.

(288) M Palmblad, S Böcker, S Degroeve, O Kohlbacher, L Käll, **WS Noble**, M Wilhelm. "Interpretation of the DOME recommendations for machine learning in proteomics and metabolomics." *Journal of Proteome Research.* 21(4):1204-1207, 2022.

(289) C Qiu, J Cao, T Li, S Srivatsan, X Huang, D Calderon, **WS Noble**, CM Disteche, M Spielmann, CB Moens, C Trapnell, J Shendure. "Systematic reconstruction of the cellular trajectories of mammalian embryogenesis." *Nature Genetics.* 54:328–341, 2022.

(290) <u>W Bittremieux</u>, <u>DH May</u>, JA Bilmes, and **WS Noble**. "A learned embedding for efficient joint analysis of millions of mass spectra." *Nature Methods.* 19(6):675–678, 2022.

(291) S Whalen*, <u>JM Schreiber</u>*, **WS Noble**, KS Pollard. "Navigating the pitfalls of applying machine learning in genomics." *Nature Reviews Genetics.* 23(3):169–181, 2022.

(292) <u>R Zhang</u>, L Meng-Papaxanthos, JP Vert, **WS Noble**. "Semi-supervised single-cell cross-modality translation using Polarbear." *Proceedings of the International Conference on Research in Computational Biology (RECOMB).* 2022.

(293) LR Heil, <u>WE Fondrie</u>, CD McGann, AJ Federation, **WS Noble**, MJ MacCoss, U Keich. "Building spectral libraries from narrow window data independent acquisition mass spectrometry data. *Journal of Proteome Research.* 21(6):1382–1391, 2022.

(294) K Mazooji, S Kannan, **WS Noble**, I Shomorony. "Fundamental limits of multi-sample flow graph decomposition." *Proceedings of the IEEE International Symposium on Information Theory.* 2403-2408, 2022.

(295) <u>M Yilmaz</u>, WE Fondrie, W Bittremieux, S Oh, **WS Noble**. "De novo mass spectrometry peptide sequencing with a transformer model." *Proceedings of the International Conference on Machine Learning.* 162:25514–25522, 2022.

(296) <u>AB Dincer</u>, <u>YY Lu</u>, D Schweppe, S Oh, **WS Noble**. "Reducing peptide sequence bias in quantitative mass spectrometry data with machine learning." *Journal of Proteome Research.* 21(7):1771–1782, 2022.

(297) B Hristov, JA Bilmes, **WS Noble**. "Linking cells across single-cell modalities by synergistic matching of neighborhood structure." *Bioinformatics (Proceedings of the European Conference on Computational Biology)*. 8(Suppl_2):ii148–ii154, 2022.

(298) J Freestone, T Short, **WS Noble**, U Keich. "Group-walk, a rigorous approach to separate false discovery rate analysis by target-decoy competition." *Bioinformatics (Proceedings of the European Conference on Computational Biology*. 38(Supplement_2):ii82–ii88, 2022.

(299) A Lin, T Short, **WS Noble**, U Keich. "Improving peptide-level mass spectrometry analysis via double competition." *Journal of Proteome Research*. 21(10):2412–2420, 2022.

(300) R Zhang, L Meng-Papaxanthos, JP Vert, **WS Noble**. "Multimodal single-cell translation and alignment with semi-supervised learning." *Journal of Computational Biology*. 29(11):1198–1212, 2022.

(301) KG Alavattam, KA Mitzelfelt, G Bonora, PA Fields, X Yang, HS Chiu, L Pabon, A Bertero, NJ Palpant, **WS Noble**, CE Murry. "Dynamic chromatin organization and regulatory interactions in human endothelial cell differentiation." *Stem Cell Reports*. 18(1):159–174, 2023.

(302) S Hasam, K Emery, **WS Noble**, U Keich. "A pipeline for peptide detection using multiple decoys." *Methods in Molecular Biology*. 2426:25–34, 2023.

(303) N Varoquaux, **WS Noble**, JP Vert. "Inference of genome 3D architecture by modeling overdispersion of Hi-C data." *Bioinformatics*. 39(1):btac838, 2023.

(304) **WS Noble**. "Ten simple rules for defining a computational biology project." *PLOS Computational Biology*. 19(1):e1010786, 2023.

(305) A Lin, BLD Kaiser, JR Hutchison, JA Bilmes, **WS Noble**. "MS1Connect: a mass spectrometry run similarity measure." *Bioinformatics*. 39(2):btad058, 2023.

(306) FLNA Acquaye, A Kertesz-Farkas, **WS Noble**. "Efficient indexing of peptides for database search using Tide." *Journal of Proteome Research*. 22(2):577–584, 2023.

(307) A Kertesz-Farkas, FLNA Acquaye, K Bimani, JK Eng, WE Fondrie, CE Grant, MR Hoopmann, A Lin, Y Lu, RL Moritz, MJ MacCoss, **WS Noble**. "The Crux toolkit for analysis of bottom-up tandem mass spectrometry proteomics data." *Journal of Proteome Research*. 22(2):561–569, 2023.

(308) D Luo, K Emery, Y He, **WS Noble**, U Keich. "Competition-based control of the false discovery proportion." *Biometrics*. 79(4):3472-3483, 2023.

(309) J Rozowsky, J Drenkow, YT Yang, G Gursoy, T Galeev, B Borsari, CB Epstein, K Xiong, J Xu, J Gao, K Yu, A Berthel, Z Chen, F Navarro, J Liu, MS Sun, J Wright, J Chang, CJF Cameron, N Shoresh, E Gaskell, J Adrian, S Aganezov, G Balderrama-Gutierrez, S Banskota, GB Corona, S Chee, SB Chhetri, GCC Martins, C Danyko, CA Davis, D Farid, NP Farrell, I Gabdank, Y Gofin, DU Gorkin, M Gu, V Hecht, BC Hitz, R Issner, M Kirsche, X Kong, BR Lam, S Li, B Li, T Li, X Li, K Zin Lin, R Luo, M Mackiewicz, JE Moore, J Mudge, N Nelson, C Nusbaum, I Popov, HE Pratt, Y Qiu, S Ramakrishnan, J Raymond, L Salichos, A Scavelli, JM Schreiber, FJ Sedlazeck, LH See, RM Sherman, X Shi, M Shi, CA Sloan, JS Strattan, Z Tan, FY Tanaka, A Vlasova, J Wang, J Werner, B Williams, M Xu, C Yan, L Yu, C Zaleski, J Zhang, JM Cherry, EM Mendenhall, **WS Noble**, Z Weng, ME Levine, A Dobin, B Wold, A Mortazavi, B Ren, J Gillis, RM Myers, MP Snyder, J Choudhary, A Milosavljevic, MC Schatz, R Guigó, BE Bernstein, TR Gingeras, M Gerstein. "The EN-TEx resource of multi-tissue personal epigenomes and variant-impact models." *Cell*. 186(7):P1493–1511, 2023.

(310) JM Schreiber, C Boix, JW Lee, H Li, Y Guan, C-C Chang, J-C Chang, A Hawkins-Hooker, B Schölkopf, G Schweikert, MR Carulla, A Canakoglu, F Guzzo, L Nanni, M Masseroli, MJ Carman, P Pinoli, C Hong, KY Yip, JP Spence, SS Batra, YS Song, S Mahony, Z Zhang, W Tan, Y Shen, Y Sun, M Shi, J Adrian, R Sandstrom, NP Farrell, J Halow, K Lee, L Jiang, X Yang, CB Epstein, JS Strattan, MP Snyder, M Kellis, **WS Noble**, A Kundaje, ENCODE Imputation Challenge Participants. "The ENCODE Imputation Challenge: A critical assessment of methods for cross-cell type imputation of epigenomic profiles." *Genome Biology*. 24:79, 2023.

(311) H Gu, H Harris, M Olshansky, Y Eliaz, A Krishna, A Kalluchi, <u>M Jacobs</u>, <u>G Cauer</u>, M Pham, SSP Rao, O Dudchenko, A Omer, K Mohajeri, S Kim, MH Nichols, ES Davis, D Udupa, AP Aiden, VG Corces, DH Phanstiel, **WS Noble**, J-S Seo, ME Talkowski, EL Aiden, MJ Rowley. "Chromatin alternates between A and B compartments at kilobase scale for subgenic organization." *Nature Communications.* 14:3033, 2023.

(312) <u>A Min</u>, TJ Durham, L Gevirtzman, **WS Noble**. "Matrix prior for data transfer between single cell data types in latent Dirichlet allocation." *PLOS Computational Biology.* 19(5):e1011049, 2023.

(313) R Yang, A Das, VR Gao, A Karbalayghareh, **WS Noble**, JA Bilmes, CS Leslie. "Epiphany: predicting Hi-C contact maps from 1D epigenomic signals." *Genome Biology.* 24:134, 2023.

(314) A Ebadi, J Freestone, **WS Noble**, U Keich. "Bridging the false discovery gap." *Journal of Proteome Research.*

(315) B Baur, J Shin, <u>JM Schreiber</u>, S Zhang, Y Zhang, M Manjunath, JS Song, **WS Noble**, S Roy. "Leveraging epigenomes and three-dimensional genome organization for interpreting regulatory variation." *PLOS Computational Biology.* 19(7):e1011286, 2023.

(316) J Dekker, F Alber, S Aufmkolk, BJ Beliveau, BG Bruneau, AS Belmont, L Bintu, A Boettiger, R Calandrelli, CM Disteche, DM Gilbert, T Gregor, AS Hansen, B Huang, D Huangfu, R Kalhor, CS Leslie, W Li, Y Li, J Ma, **WS Noble**, PJ Park, JE Phillips-Cremins, KS Pollard, S Rafelski, B Ren, Y Ruan, Y Shav-Tal, Y Shen, J Shendure, X Shu, C Strambio-De-Castillia, A Vertii, H Zhang, S Zhong. "Spatial and temporal organization of the genome: current state and future aims of the 4D Nucleome Project." *Molecular Cell.* 83(15):2624–2640, 2023.

(317) L Meng-Papaxanthos, <u>R Zhang</u>, <u>G Li</u>, M Cuturi, **WS Noble**, JP Vert. "LSMMD-MA: Scaling multimodal data integration for single-cell genomics data analysis." *Bioinformatics.* 39(7):btad420, 2023.

(318) <u>L Harris</u>, W Fondrie, S Oh, **WS Noble**. "Evaluating proteomics imputation methods with improved criteria." *Journal of Proteome Research.* 22(11):3427–3438, 2023.

(319) D Mar, I Babenko, <u>R Zhang</u>, **WS Noble**, O Denisenko, T Vaisar, K Bomsztyk. "A high throughput PIXUL-Matrix-based toolbox to profile frozen and FFPE tissues multiomes." *Laboratory Investigation.* 104(1):100282, 2024. `https://www.biorxiv.org/content/10.1101/2023.03.16.533031v1`

(320) <u>R Aguilar</u>, CK Camplisson, Q Lin, KH Miga, **WS Noble**, BJ Beliveau. "Tigerfish designs oligonucleotide-based in situ hybridization probes targeting intervals of highly repetitive DNA at the scale of genomes." *Nature Communications.* 15:1027, 2024.

(321) T Hollin, S Abel, C Banks, <u>B Hristov</u>, J Prudhomme, <u>K Hales</u>, L Florens, **WS Noble**, KG Le Roch. "Proteome-Wide Identification of RNA-dependent proteins and an emerging role for RNAs in *Plasmodium falciparum* protein complexes." *Nature Communications.* In press.

(322) S Zheng, N Thakkar, H Harris, M Zhang, S Liu, M Gerstein, EL Aiden, MJ Rowley, **WS Noble**, G Gursoy, R Singh. "Predicting A/B compartments from histone modifications using deep learning." *ACM-BCB '23: Proceedings of the 14th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics.* In press.

(323) J Freestone, **WS Noble**, U Keich. "Analysis of tandem mass spectrometry data with CONGA: Combining Open and Narrow searches with Group-wise Analysis." *Journal of Proteome Research.* In press. `https://www.biorxiv.org/content/10.1101/2023.05.02.539167v1`

(324) A Lin, <u>D See</u>, WE Fondrie, U Keich, **WS Noble**. "Target-decoy false discovery rate estimation using Crema." *Proteomics.* In press. `https://www.biorxiv.org/content/10.1101/2023.06.18.545038v1`

(325) J Freestone, L Käll, **WS Noble**, U Keich. "Semi-supervised learning while controlling the FDR with an application to tandem mass spectrometry analysis." *Proceedings of the International Conference on Research in Computational Biology (RECOMB).* 2024.

(326) T Hollin, S Abel, C Banks, <u>B Hristov</u>, J Prudhomme, <u>K Hales</u>, L Florens, **WS Noble**, KG Le Roch. "Proteome-wide identification of RNA-dependent proteins: an emerging role for RNAs in *Plasmodium falciparum* protein complexes." *Nature Communications*. 15:1365, 2024.

(327) C Qiu, BK Martin, IC Welsh, RM Daza, TM Le, X Huang, EK Nichols, ML Taylor, O Fulton, DR O'Day, AR Gomes, S Ilcisin, S Srivatsan, X Deng, CM Disteche, **WS Noble**, N Hamazaki, CB Moens, D Kimelman, J Cao, AF Schier, M Spielmann, SA Murray, C Trapnell, J Shendure. "A single-cell time-lapse of mouse prenatal development from gastrula to birth." *Nature*. 626:1084–1093, 2024.

(328) T Fang, Y Liu, A Woicik, M Lu, <u>A Jha</u>, <u>X Wang</u>, <u>G Li</u>, <u>B Hristov</u>, Z Liu, H Xu, **WS Noble**, S Wang. "Enhancing Hi-C contact matrices for loop detection with Capricorn, a multi-view diffusion model." *Bioinformatics (Proceedings of the ISMB)*. In press. `https://www.biorxiv.org/content/10.1101/2023.10.25.564065v1`

(329) <u>V Ananth</u>*, <u>J Sanders</u>*, <u>M Yilmaz</u>, S Oh and **WS Noble**. "A learned score function improves the power of mass spectrometry database search." *Bioinformatics (Proceedings of the ISMB)*. In press. `https://www.biorxiv.org/content/10.1101/2024.01.26.577425v1`

Additional Publications

(330) **WN Grundy**. *Making a Thinking Robot*. Senior honors thesis, Symbolic Systems Program, Stanford University, 1991. Prof. Fred Dretske, advisor.

(331) **WN Grundy**. "Building a simple recognition system using HTK." Technical report, Entropic Research Laboratory, Washington, DC. August, 1994.

(332) **WN Grundy**. *A Bayesian Approach to Motif-based Protein Modeling*. Ph.D. dissertation. University of California, San Diego. June, 1998. Prof. Charles Elkan, advisor.

(333) TL Bailey, ME Baker, CP Elkan and **WN Grundy**. "MEME, MAST, and Meta-MEME: New Tools for Motif Discovery in Protein Sequences" in *Pattern Discovery in Biomolecular Data: Tools, Techniques and Applications*. J Wang, B Shapiro and D Shasha, ed. Oxford UP, 1999.

(334) **WS Noble**. "Support vector machine applications in computational biology." *Kernel Methods in Computational Biology*. B Schölkopf, K Tsuda and JP Vert, ed. MIT Press, 2004. pp. 71–92.

(335) **WS Noble**. "Support vector machine software." *Encyclopedia of Genomics, Proteomics and Bioinformatics*. M Dunn, L Jorde, P Little and S Subramaniam, ed. Wiley, 2004.

(336) **WS Noble**. "Data hoarding is harming proteomics." *Nature Biotechnology*. 22(10):1209, 2004.

(337) **WS Noble**, R Kuang, C Leslie and J Weston. "Identifying remote protein homologs by network propagation." *FEBS Journal*. 272(20):5119–5128, 2005.

(338) J Weston, C Leslie, E Ie and **WS Noble**. "Semi-supervised protein classification using cluster kernels." *Semi-supervised Learning*. O Chapelle, B Schölkopf and A Zien, ed. MIT Press, 2005. pp. 321–338.

(339) **WS Noble** and A Ben-Hur. "Integrating information for protein function prediction." *Bioinformatics: From Genomes to Therapies*. Ed. by T Lengauer. Wiley, 2006.

(340) **WS Noble.** "What is a support vector machine?" *Nature Biotechnology*. 24(12):1565–1567, 2006.

(341) J Weston, G Bakir, O Bousquet, <u>T Mann</u>, **WS Noble** and B Schölkopf. "Joint kernel maps." *Predicting Structured Data*. Ed. by G Bakir, T Hoffmann, B Schölkopf, AJ Smola, B Taskar and VSN Vishwanathan. MIT Press, 2007.

(342) L Käll, JD Storey, MJ MacCoss and **WS Noble**. "Posterior error probability and false discovery rates: Two sides of the same coin." *Journal of Proteome Research*. 7(1):40–44, 2008. PMID18052118.

(343) **WS Noble** and C Leslie. "Learning models of biological sequences." In *Encyclopedia of Machine Learning.* Ed. by C Sammut and GI Webb. Springer Verlag, 2011.

(344) Y Qi and **WS Noble**. "Protein interaction networks: protein domain interaction and protein function prediction." In *Handbook of Statistical Bioinformatics.* Ed. by HH Lu, B Schölkopf and H. Zhao. Springer Verlag, 2011.

(345) **WS Noble**, CA Blau, J Dekker, ZJ Duan and Y Mao. "The structure and function of chromatin and chromosomes." *Proceedings of the Pacific Symposium on Biocomputing.* 17:434–440, 2012.

(346) OR Serang and **WS Noble**. "A review of statistical methods for protein identification using tandem mass spectrometry." *Statistics and its Interface.* 5:3–20, 2012.

(347) **WS Noble** and MJ MacCoss. "Computational and statistical analysis of protein mass spectrometry data." *PLOS Computational Biology.* 8(1):e1002296, 2012.

(348) BJ Diament, MJ MacCoss and **WS Noble**. "On the feasibility and utility of exploiting real time database search to improve adaptive peak selection." arXiv:1207.5848v1, 2012.

(349) V Granholm, **WS Noble** and L Käll. "A cross-validation scheme for machine learning algorithms in shotgun proteomics." *BMC Bioinformatics.* 13(Suppl 16):S3, 2012.

(350) Z Duan, M Andronescu, K Schutz, C Lee, J Shendure, S Fields, **WS Noble** and CA Blau. "A genome-wide 3C-method for characterizing the three-dimensional architectures of genomes." *Methods.* 58(3):277-288, 2012.

(351) GG Yardımcı, **WS Noble**. "A predictive model of 3D domain formation via CTCF-mediated extrusion." *Proceedings of the National Academy of Sciences.* 112(47):14404–14405, 2015.

(352) A Hu, **WS Noble** and A Wolf-Yadlin. "Technical advances in proteomics: new developments in data independent acquisition analytical software." *F1000 Faculty Reviews.* 5(F1000 Faculty Rev):419, 2016.

(353) A van der Velde, M Purcaro, **WS Noble**, and Z Weng. "LR-DNase: Predicting TF binding prediction from DNase-seq data." `http://biorxiv.org/content/early/2016/10/24/082594`, 2016.

(354) N Hiranuma, J Liu, C Song, J Goldsmith, MO Dorschner, CC Pritchard, KA Burton, EM Mahen, S Blau, FM Senecal, WL Monsky, S Parker, SC Schmechel, SK Allison, VK Gadi, SR Salama, AJ Radenbaugh, M Goldman, JM Johnsen, S Heimfeld, V Komashko, M LaMadrid-Hermannsfeldt, Z Duan, SC Benz, P Soon-Shiong, D Haussler, J Zhu, WL Ruzzo, **WS Noble**, and CA Blau. "Cis-compound mutations are prevalent in triple negative breast cancer and can drive tumor progression." `http://biorxiv.org/content/early/2016/11/05/085316`, 2016.

(355) J Schreiber, MW Libbrecht, J Bilmes, and **WS Noble**. "Nucleotide sequence and DNaseI sensitivity are predictive of 3D chromatin architecture." `https://www.biorxiv.org/content/early/2018/07/10/103614`, 2017.

(356) J Schreiber, YY Lu, **WS Noble**. "Ledidi: Designing genomic edits that induce functional activity." `https://www.biorxiv.org/content/10.1101/2020.05.21.109686v1`, 2020.

(357) W Bittremieux, **WS Noble**. "A neural network for large-scale clustering of peptide mass spectra." *Nature Methods.* 19(6):658–659, 2022.

(358) S Hasam, K Emery, **WS Noble**, U Keich. "A pipeline for peptide detection using multiple decoys." *Statistical Analysis of Proteomic Data.* T. Burger, ed. *Methods in Molecular Biology.* 2426:25–34, 2022.

(359) A Sur, **WS Noble**, S Sullivan, P Myler. "Edison: measuring scaffolding accuracy with edit distance." *bioRxiv.* `https://www.biorxiv.org/content/10.1101/2022.03.25.484952v1`

(360) A Sur, **WS Noble**, PJ Myler. "A benchmark of Hi-C scaffolders using reference genomes and de novo assemblies." *bioRxiv.* `https://www.biorxiv.org/content/10.1101/2022.04.20.488415v1`

Submitted for Publication

(361) V Agarwal, F Inoue, M Schubach, BK Martin, PM Dash, Z Zhang, A Sohota, **WS Noble**, <u>GG Yardimci</u>, M Kircher, J Shendure, N Ahituv. "Massively parallel characterization of transcriptional regulatory elements in three diverse human cell types." *bioRxiv.*

(362) RA Rodriguez-Mias, KN Hess, BY Ruiz, IR Smith, AS Barente, SM Zimmerman, <u>YY Lu</u>, **WS Noble**, S Fields, J Villén. "Proteome-wide identification of amino acid substitutions deleterious for protein function." *bioRxiv.* `https://www.biorxiv.org/content/10.1101/2022.04.06.487405v1`

(363) <u>M Yilmaz</u>*, WE Fondrie*, W Bittremieux*, <u>R Nelson</u>, <u>V Ananth</u>, S Oh, **WS Noble**. "Sequence-to-sequence translation from mass spectra to peptides with a transformer model." `https://www.biorxiv.org/content/10.1101/2023.01.03.522621v1`

(364) A Ebadi, D Luo, J Freestone, **WS Noble**, U Keich. "Bounding the FDP in competition-based control of the FDR." `https://arxiv.org/abs/2302.11837`

(365) <u>A Jha</u>, SC Bohaczuk, Y Mao, J Ranchalis, BJ Mallory, <u>AT Min</u>, MO Hamm, E Swanson, C Finkbeiner, T Li, D Whittington, **WS Noble**, AB Stergachis, MR Vollger. "Fibertools: fast and accurate DNA-m6A calling using single-molecule long-read sequencing." *bioRxiv* `https://www.biorxiv.org/content/10.1101/2023.04.20.537673v1`

(366) J Freestone, **WS Noble**, U Keich. "Re-investigating the correctness of decoy-based false discovery rate control in proteomics tandem mass spectrometry." *bioRxiv.* `https://www.biorxiv.org/content/10.1101/2023.06.21.546013v1`

(367) H Fang, AR Tronco, <u>G Bonora</u>, T Nguyen, J Thakur, JB Berletch, GN Filippova, S Henikoff, J Shendure, **WS Noble**, CM Disteche, X Deng. "CTCF-mediated insulation and chromatin environment modulate Car5b escape from X inactivation" *bioRxiv.* `https://www.biorxiv.org/content/10.1101/2023.05.04.539469v1`

(368) <u>B Hristov</u>, **WS Noble**, A Bertero. "Systematic identification of inter-chromosomal interaction networks supports the existence of specialized RNA factories." `https://www.biorxiv.org/content/10.1101/2023.09.21.558852v1`

(369) <u>W Chen</u>, **WS Noble**, <u>Y Lu</u>. "DeepROCK: Error-controlled interaction detection in deep neural networks."

(370) R Zhang, M Yang, J Schreiber, DR O'Day, J Turner, J Shendure, CM Disteche, X Deng, **WS Noble**. "Cross-species imputation and comparison of single-cell transcriptomic profiles." `https://www.biorxiv.org/content/10.1101/2023.10.19.563173v1`

(371) Y Lu, **WS Noble**, U Keich. "A BLAST from the past: revisiting BLAST's E-value."

(372) J Freestone, L Käll, **WS Noble**, U Keich. "Semi-supervised learning while controlling the FDR with an application to tandem mass spectrometry analysis." `https://www.biorxiv.org/content/10.1101/2023.10.26.564068v3`

(373) <u>G Li</u>, H-J Kim, S Pendyala, <u>R Zhang</u>, CM Disteche, J-P Vert, X Deng, D Fowler, **WS Noble**. "Pseudotime analysis for time-series single-cell sequencing and imaging data." `https://www.biorxiv.org/content/10.1101/2023.11.03.565575v1`

(374) CC Wu, KA Tsantilas, J Park, D Plubell, <u>JA Sanders</u>, P Naicker, I Govender, S Buthelezi, S Stoychev, J Jordaan, G Merrihew, E Huang, ED Parker, M Riffle, AN Hoofnagle, **WS Noble**, KL Poston, TJ Montine, MJ MacCoss. "Mag-Net: Rapid enrichment of membrane-bound particles enables high coverage quantitative analysis of the plasma proteome."

(375) A Min, JM Schreiber, A Kundaje, **WS Noble**. "Predicting chromatin conformation contact maps."
https://www.biorxiv.org/content/10.1101/2024.04.12.589240v1

Invited lectures (since 2015)

"Deep learning applications in proteomics mass spectrometry." Deep Learning Affinity Group, Fred Hutch, May 2, 2023.

"Deep learning applications in mass spectrometry proteomics and single-cell genomics." Keynote, 9th International School on Deep Learning, Bari, Italy, April 3–7, 2023.

"Deep learning applications in proteomics mass spectrometry." AIC Laboratory, HSE University, Oct 13, 2022.

"A machine learning model for translation between modalities of single-cell sequencing data." CHARGE Consortium annual meeting, Seattle, WA, Oct 12, 2022.

"Deep learning applications in proteomics mass spectrometry." Kipoi Machine Learning seminar, Oct 5, 2022.

"Learning embeddings of bulk and single-cell genomic data for imputation and multi-omic integration." Keynote, Omics Research Day, Simon Fraser University, Feb 8, 2020.

"Machine learning methods for making sense of big genomic and proteomic data." Keynote, Machine Learning for Computational Biology Workshop, Vancouver, BC, Dec 14, 2019.

"Imputing and prioritizing epigenomics experiments." Cancer Epigenetics Affinity Group, Fred Hutch, Sep 3, 2019.

"NOT all the things: submodular representative set selection for when big data is too big." Keynote, Machine Learning in Computational and Systems Biology subgroup, Intelligent Systems for Molecular Biology Conference, Basel, Switzerland, Jul 25, 2019.

"Traveling across spaces: the power of embedding genomic and proteomic data into a latent space." Keynote, Intelligent Systems for Molecular Biology Conference, Basel, Switzerland, Jul 22, 2019.

"Learning latent embeddings of genomic and proteomic data." Genetics, Bioinformatics and Systems Biology Colloquium, UC San Diego, Mar 8, 2019.

"Learning latent embeddings of genomic and proteomic data." DCMB Seminar, University of Michigan, Feb 27, 2019.

"Unsupervised embedding of single-cell Hi-C data." Joint Statistical Meetings, Vancouver, BC. Jul 20, 2018.

"Modeling and predicting the 3D genome." Keynote, 3D Genomics Special Interest Group, Intelligent Systems for Molecular Biology, Chicago, IL. Jul 7, 2018.

"Modeling the 3D architecture of the genome." Keystone Symposium on Chromatin Architecture and Chromosome Organization, Whistler, BC, Canada. Mar 27, 2018.

"Machine learning methods for making sense of big genomic data." Computational Genomics Winter Institute, UCLA, March 1, 2018.

"Machine learning applications in genetics and genomics." Data Science Seminar, Fred Hutch, Jan 16, 2018.

"Machine learning and statistical challenges in protein mass spectrometry." Huck Institute Distinguished Lecture Series, Penn State University, Nov 7, 2017.

"Machine learning and statistical challenges in protein mass spectrometry." Penn Bioinformatics Forum, University of Pennsylvania, Nov 8, 2017.

"Machine learning and statistical challenges in protein mass spectrometry." Keynote speaker, X-Meeting (13th International Conference of the Brazilian Bioinformatics and Computational Biology Association), Sao

Paulo, Brazil, Oct 4-6, 2017.

"Machine learning methods for annotating and extending big, heterogeneous genomic data." Distinguished Lecture Series, Department of Computer Science, University of Illinois at Urbana-Champaign, Nov 14, 2016.

"Chromosome organization and chromatin dynamics during the transmission stages of Plasmodium parasites." Monday Seminar Series, Center for Infectious Disease Research, Seattle, WA. Nov 7, 2016.

"Machine learning methods for annotating and extending big, heterogeneous genomic data." Genomics Seminar Series, University of Wisconsin, Madison, WA. Oct 27, 2016.

"Machine learning analysis of big, heterogeneous genomic data." Machine Learning Seminar, University of Washington. Oct. 18, 2016.

"Joint imputation of epigenomics data by three dimensional tensor factorization." BigLS 2016 ACM International Workshop on Big Data in Life Sciences, Oct 2, 2016.

"Selecting genomics assays." Workshop on Regulatory Genomics and Epigenomics, Simons Institute, UC Berkeley, Berkeley, CA. Mar. 10, 2016.

"Selecting genomics assays and making sense of the resulting data." Genomics@JHU seminar series. Johns Hopkins University, Baltimore, MD. Feb 23, 2016.

"Gene regulation in 3D." Keynote speaker, RECOMB ISCB Regulatory and Systems Genomics Conference, Philadelphia, PA. Nov 17, 2015.

"How to fit 6 billion nucleotides into a 6 micron nucleus." RECOMB Satellite Conference on Bioinformatics Education, Chevy Chase, MD. Nov 15, 2015.

"Joint annotation of chromatin state and chromatin conformation." Session on "Computational/statistical methods for integrative -omics" at the Joint Statistical Meeting, Seattle, WA. Aug 9, 2015.

"Modeling the 4D nucleome." Workshop on "Interdisciplinary Approaches to Biomedical Data Science Challenges : SAMSI Innovations Lab", Statistical and Applied Mathematical Sciences Institute, Research Triangle Park, NC. July 21, 2015.

"Gene regulation in 3D." CMO/BIRS workshop on "Rules of Protein-DNA Recognition: Computational and Experimental Advances," Oaxaca, Mexico. June 25, 2015.

"Linear and three-dimensional genome architecture." cBio Seminar, Memorial Sloan-Kettering Cancer Center. April 16, 2015.

"Three-dimensional genome architecture." Biostatistics Seminar, University of North Carolina, March 26, 2015.
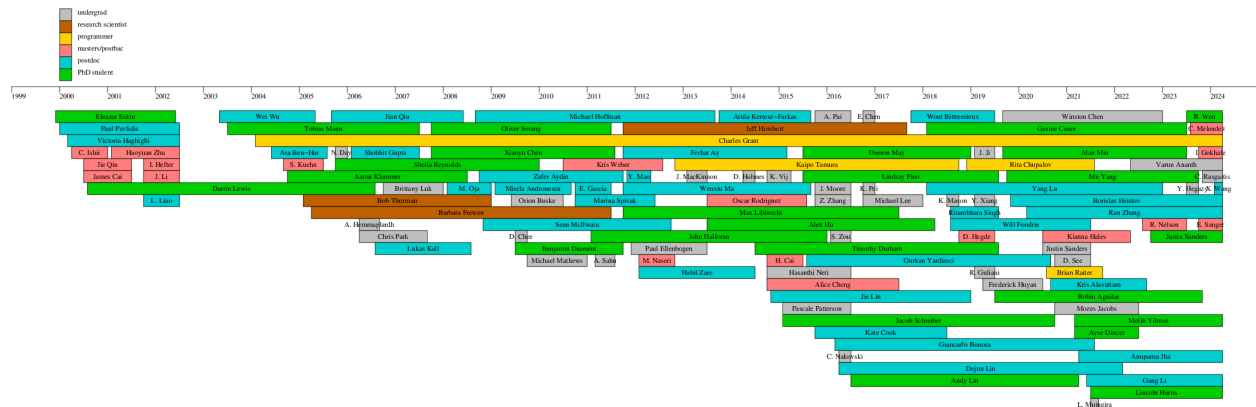
Current graduate students and postdoctoral associates

Xiao Wang, Postdoctoral Associate, Genome Sciences, University of Washington

Dates: 01/24–present
Prior Degree: Ph.D., Computer Science, Purdue University, 2023
Project: "A foundation model for Hi-C data"
Funding: R01 HL169156

Gang Li, Postdoctoral Associate, Genome Sciences, University of Washington
Dates: 05/21–present
Prior Degree: Ph.D., Interdisciplinary Statistics and Operations Research, UNC Chapel Hill, 2021
Project: "Integration of single-cell genomics data and microscopy images in the context of visual cell sorting"
Funding: NIH U54 DK107979

Anupama Jha, Postdoctoral Associate, Genome Sciences, University of Washington
Dates: 03/21–present
Prior Degree: Ph.D., Computer and Information Science, University of Pennsylvania, 2021
Project: "Cross-species prediction of chromatin accessiblity and 3D genome architecture data"
Funding: NIH U54 DK107979

Ran Zhang, Postdoctoral Associate, Genome Sciences, University of Washington
Dates: 02/20–present
Prior Degree: Ph.D., Department of Computer Science, Princeton University, 2019
Project: "Integration of multiple single-cell 3D genome architecture data sets"
Funding: NIH U54 DK107979

Borislav Hristov, Postdoctoral Associate, Genome Sciences, University of Washington
Dates: 10/19–present
Prior Degree: Ph.D., Department of Computer Science, Princeton University, 2019
Project: "Analysis methods for deciphering lncRNA-DNA binding in *Plasmodium falciparum*"
Funding: NIH R01 AI136511

Justin Sanders, Ph.D. student, Paul G. Allen School of Computer Science and Engineering
Dates: 09/22–present
Prior degree: B.S., Computer Science, Brown University
Project: Target-decoy training of a neural network search engine for mass spectra
Funding: NSF 2245300

Melih Yilmaz, Ph.D. student, Paul G. Allen School of Computer Science and Engineering
Dates: 02/21–present
Prior degree: B.S., Electrical and Electronic Engineering, Koc University, Istanbul, Turkey
Project: *De novo* spectrum identification using deep neural networks
Funding: NSF 2245300

Lincoln Harris, Ph.D. student, Department of Genome Sciences
Dates: 06/21–present
Prior degree: B.A., Biology, Swarthmore College, 2017
Project: Deep tensor-based imputation of mass spectrometry data
Funding: NSF 2245300

Former graduate students and postdoctoral associates

Robin Aguilar, Ph.D. student, Genome Sciences, University of Washington
Dates: 06/19–10/23
Prior degree: B.S., Biochemistry, De Pauw University, 2018
Project: Oligo probe design for high-throughput fluorescence microscopy
Current position: Scientific and Marketing Communications Specialist, A-Alpha Bio

Mu Yang, Ph.D. student, Biomedical Informatics and Medical Education, University of Washington
Dates: 09/19–09/23
Prior degree: B.S., Public Health, National Taiwan University, 2015
Project: Cross-species analysis of single-cell RNA-seq data
Current position: unknown

Gesine Cauer, Ph.D. student, Genome Sciences, University of Washington
Dates: 01/18–06/23
Prior degree: B.S., Biology, Haverford College, 2012
Project: "Inference of 3D structure of diploid genomes from Hi-C data"
Current position: medical leave

Alan Min, Ph.D. student, Statistics, University of Washington
Dates: 08/19–06/23
Prior degree: B.S., Computer science, math, and statistics, Purdue University, 2018
Project: Latent Dirichlet allocation modeling of single-cell expression and chromatin accessibility data
Current position: Medical school, University of Washington

Yang Lu, Postdoctoral Associate, Genome Sciences, University of Washington
Dates: 01/18–present
Prior Degree: Ph.D., Department of Computer Science, University of Southern California, 2017
Project: "Detecting linked peptide pairs in data-independent acquisition tandem mass spectrometry data."
Current position: Assistant Professor, Cheriton School of Computer Science, University of Waterloo

Kris Alavattam, Postdoctoral Associate, Genome Sciences, University of Washington
Dates: 08/20–present
Prior Degree: Ph.D., Cancer and Cell Biology, University of Cincinnati, 2020
Project: "Chromatin architecture of human epithelial cells"
Current position: United States Peace Corps Volunteer

Kianna Hales, Masters student, Department of Genome Sciences
Dates: 06/20–03/22
Prior degree: B.S., Molecular, Cellular & Developmental Biology / Data Science, University of Washington, 2019
Project: Deep neural network de-noising of tandem mass spectrometry data
Current position: Data analyst, SEPHORA

Ayse Dincer, Ph.D. student, Paul G. Allen School of Computer Science and Engineering
Dates: 02/21–06/22
Prior degree: B.S., Computer Science, Bilkent University, Ankarra, Turkey, 2017
Project: Correcting peptide-specific biases in quantitative mass spectrometry data
Current position: Uber Research

Dejun Lin, Postdoctoral Associate, Genome Sciences, University of Washington
Dates: 03/16–02/22
Prior Degree: Ph.D., Biophysics, University of Rochester, 2015
Project: "Molecular dynamics modeling of 3D genome architecture"
Current position: unknown

Giancarlo Bonora, Postdoctoral Associate, Genome Sciences, University of Washington
Dates: 02/16–07/21
Prior Degree: Ph.D., Computer Science, University of California, Los Angeles, 2015
Project: "Computational methods for analysis of 3D nucleome data"
Current position: Senior Bioinformatics Scientist, R&D department, Predicine

Andy Lin, Ph.D. student, Genome Sciences, University of Washington
Dates: 06/16–03/21
Prior degree: B.S., Cellular and Molecular Biology and Bioinformatics, University of Michigan, 2012

Project: "Unsupervised ranking of mass spectrometry runs"
Current position: Linus Pauling Distinguished Postdoctoral Fellow, Pacific Northwest National Labs

Jacob Schreiber, Ph.D. student, Computer Science and Engineering, University of Washington
Dates: 01/15–09/20
Prior degree: B.S., Biomolecular Engineering, University of California, Santa Cruz, 2013
Project: "Deep tensor factorization for large-scale imputation of epigenomic data"
Current position: Postdoctoral Fellow with Prof. Anshul Kundaje, Department of Genetics, Stanford University

Gürkan Yardımcı, Postdoctoral Associate, Genome Sciences, University of Washington
Dates: 07/15–08/20
Prior Degree: Ph.D., Computational Biology and Bioinformatics, Duke University, 2014
Project: "Computational methods for analysis of 3D nucleome data"
Current position: Assistant Professor, Oncological Sciences Division, Oregon Health Sciences University

Timothy Durham, Ph.D. student, Genome Sciences, University of Washington
Dates: 06/14–08/19
Prior degree: B.S., Biology and Computer Science, Williams College, 2009
Project: "Machine learning methods to impute epigenomic maps"
Current position: Scientist, Broad Institute

Ritambhara Singh, Postdoctoral Associate, Genome Sciences, University of Washington
Date: 07/18–07/19
Prior Degree: Ph.D., Department of Computer Science, University of Virginia, 2018
Project: "A deep neural network approach to improving resolution of single-cell Hi-C data."
Current position: Assistant Professor, Department of Computer Science, Brown University

Lindsay Pino, Ph.D. student, Genome Sciences, University of Washington
Dates: 06/15–06/19
Prior degree: B.S., Biochemistry and Molecular Biology, Pennsylvania State University, 2009
Project: "Proteomics technologies to uncover the molecular phenotypes of aging"
Current position: Postdoctoral fellow, Ben Garcia lab, University of Pennsylvania.

Wout Bittremieux, Postdoctoral Associate, Genome Sciences, University of Washington
Dates: 09/17–06/19
Prior Degree: Ph.D., Computer Science, University of Antwerp, 2017
Project: "Generating theoretical peptide mass spectra using a generative adversarial network."
Current position: Postdoctoral fellow with Prof. Pieter Dorrestein, Skaggs School of Pharmacy and Pharmaceutical Sciences, UC San Diego

Jie Liu, Postdoctoral Associate, Genome Sciences, University of Washington
Dates: 10/14–12/18
Prior Degree: Ph.D., Computer Science, University of Wisconsin–Madison, 2014
Project: "Summarization of big cancer data"
Current position: Assistant Professor, Department of Computational Medicine and Bioinformatics, University of Michigan

Damon May, Ph.D. student, Genome Sciences, University of Washington
Dates: 06/15–12/18
Prior degree: M.S., Computer Science and Engineering, University of Washington, 2005
Project: "A learned embedding for efficient joint analysis of millions of mass spectra"
Current position: Computational Immunologist, Adaptive Biotechnologies

Kate Cook, Postdoctoral Associate, Genome Sciences, University of Washington
Dates: 09/15–06/18
Prior Degree: Ph.D., Molecular Genetics, University of Toronto, 2015
Project: "Relating gene expression to genome architecture in *Plasmodium falciparam*"

Current position: Research Associate, Ottawa Hospital Research Institute, University of Ottawa

Alex Hu, Ph.D. student, Genome Sciences, University of Washington
Dates: 06/13–03/18
Prior degree: B.S., Biomedical Engineering and B.S., Biology, UT Austin, 2012
Project: "Computational methods for the analysis of data-independent acquisition shotgun proteomics data."
Current position: Research scientist, Benaroya Research Institute

Max Libbrecht, Ph.D. student, Computer Science and Engineering, University of Washington
Dates: 09/11–06/17
Prior degree: B.S., Computer Science, Stanford University, 2011
Project: "Understanding human genome regulation through entropic graph-based regularization and submodular optimization"
Current position: Assistant Professor, Department of Computer Science, Simon Fraser University.

John Halloran, Ph.D. student, Electrical Engineering, University of Washington
Dates: 01/11–12/15
Prior degree: M.S., Electrical Engineering, University of Hawaii at Manoa, 2010
Project: "Graphical Models for Peptide Identification of Tandem Mass Spectra."
Current position: Postdoctoral Fellow, Division of Biostatistics, UC Davis

Attila Kertesz-Farkas, Postdoctoral Associate, Genome Sciences, University of Washington
Dates: 09/13–08/15
Prior degree: Ph.D., Computer Science, University of Szeged, Hungary, 2010
Project: "Accurate and rapid identification of post-translationally modified peptides from shotgun proteomics data."
Current position: Assistant Professor, School of Data Analysis and Artificial Intelligence, the Faculty of Informatics, National Research University Higher School of Economics in Moscow, Russian Federation.

Wenxiu Ma, Postdoctoral Associate, Genome Sciences, University of Washington
Dates: 09/11–08/15
Prior Degree: Ph.D., Computer Science, Stanford University, 2012
Project: "Analysis of ChIP-seq and chromatin architecture data"
Current position: Assistant Professor, Department of Statistics, University of California, Riverside.

Ferhat Ay, Postdoctoral Associate, Genome Sciences, University of Washington
Dates: 09/11–02/15
Prior Degree: Ph.D., Computer Science, University of Florida, 2011
Project: "Inference of genome architecture from DNA-DNA interaction data."
Current position: Institute Leadership Assistant Professor of Computational Biology, La Jolla Institute for Allergy and Immunology

Habil Zare, Postdoctoral Associate, Genome Sciences, University of Washington
Dates: 01/12–06/14
Prior Degree: Ph.D., Computer Science, University of British Columbia, 2011
Project: "Inferring clonal composition from multiple sections of a breast cancer"
Current position: Assistant Professor, Department of Cell Systems & Anatomy, University of Texas Health Science Center at San Antonio

Michael Hoffman, Postdoctoral Associate, Genome Sciences, University of Washington
Dates: 08/08–08/13
Prior degree: Ph.D., Biology, Trinity College, University of Cambridge, 2008
Project: "Unsupervised dynamic Bayesian networks for segmentation of heterogeneous genomic data sets."
Current position: Assistant Professor, Department of Medical Biophysics and the Department of Computer Science, University of Toronto. Scientist (Principal Investigator), Princess Margaret Cancer Center, Toronto, CA.

Sean McIlwain, Postdoctoral Associate, Genome Sciences, University of Washington

Dates: 10/08–09/12
Prior degree: Ph.D., Department of Computer Science, University of Wisconsin, Madison, 2008
Project: "Analysis of tandem mass spectra from cross-linked peptides"
Current position: Bioinformatics Researcher, Greater Lakes Bioenergy Research Center, University of Wisconsin

Kris Weber, Masters student, Computer Science and Engineering, University of Washington
Dates: 06/10–07/12
Prior degree: B.S., Computer Science and Engineering, University of Washington, 2010
Project: "Characterizing the clonal population of a single cancer using high-throughput sequencing."
Current position: Amazon.com.

Marina Spivak, Postdoctoral Associate, Genome Sciences, University of Washington
Dates: 09/10–05/12
Prior Degree: Ph.D., Computer Science, New York University, 2010
Project: "Joint peptide and protein identification from shotgun proteomics data."
Current position: Data Scientist, Center for Computational Biology, Simons Foundation.

Yi Mao, Postdoctoral Associate, Genome Sciences, University of Washington
Dates: 10/11–04/12
Prior Degree: Ph.D., Physical Chemistry, Northwestern University, 2000
Project: "Modeling mammalian genomes."
Current position: Unknown

Benjamin Diament, Ph.D. student, Computer Science and Engineering, University of Washington
Dates: 06/09–09/11
Prior degree: M.S., Computer Science, UC Berkeley
Project: "Ultrafast peptide identification from tandem mass spectra: method and applications."
Current position: Mass spectrometry software start-up

Zafer Aydin, Postdoctoral Associate, Genome Sciences, University of Washington
Dates: 09/08–09/11
Prior degree: Ph.D., Electrical and Computer Engineering, Georgia Institute of Technology, 2008
Project: "Supervised learning for analysis of *C. elegans* single cell expression images"; "Dynamic Bayesian networks for predicting protein secondary structure"
Current position: Assistant Professor, Computer Enginering Department, Abdullah Gul University, Kayseri, Turkey.

Xiaoyu Chen, Ph.D. student, Computer Science and Engineering, University of Washington
Dates: 09/07–07/11
Prior degree: M.S., Computer Science, McGill University, 2006
Project: "Discovering and characterizing tissue-specific regulatory motifs in the human genome."
Current position: Illumina, San Diego, CA.

Eric Garcia, Postdoctoral Associate, Genome Sciences, University of Washington
Dates: 09/10–06/11
Prior degree: Ph.D., Electrical Engineering, University of Washington, 2010
Project: "The relationship between structure and function of the human genome."
Current position: Machine Learning Engineer, Spotify.

Oliver Serang, Ph.D. student, Genome Sciences, University of Washington
Dates: 09/07–06/11
Prior degree: B.S., Computer Science, University of North Carolina, 2006
Project: "Accurate and efficient characterization of complex protein mixtures."
Current position: Assistant Professor, Department of Computer Science, University of Montana

Mirela Andronescu, Postdoctoral Associate, Genome Sciences, University of Washington
Dates: 01/09–08/10

Prior degree: Ph.D., Computer Science, University of British Columbia
Project: "Characterizing the 3D structure of the yeast and human genomes."
Current position: Bioinformatics Scientist, Shah Lab, University of British Columbia

Sheila Reynolds, Ph.D. student, Electrical Engineering, University of Washington
Dates: 09/05–12/09
Prior degree: M.S, Electrical Engineering, George Washington University, 1990.
Project: "Classification and labeling of biological sequences using dynamic Bayesian networks."
Current position: Senior Research Scientist, Ilya Shmulevich lab, Institute for Systems Biology.

Merja Oja, Postdoctoral Associate, Genome Sciences, University of Washington
Dates: 01/08–12/08
Prior degree: Ph.D., Department of Computer Science, University of Helsinki, 2007
Project: "Multitask learning for predicting protein properties from sequence."
Current position: Researcher, Protein Production Lab, VTT Technical Research Centre of Finland.

Lukas Käll, Postdoctoral Associate, Genome Sciences, University of Washington
Dates: 07/06–07/08
Prior degree: Ph.D., Bioinformatics, Karolinska Institute, 2006
Project: "Machine learning and statistical tools for enhanced peptide identification from shotgun proteomics data."
Current position: Professor, Applied Systems Biology, KTH—Royal Institute of Technology

Aaron Klammer, Ph.D. student, Genome Sciences, University of Washington
Dates: 09/04–06/08
Prior degree: B.A., Biology and Biochemistry, University of California, Santa Barbara, 1999
Project: "A machine learning approach to peptide identification with tandem mass spectrometry"
Current position: Pacific Biosciences, Menlo Park, CA.

Jian Qiu, Postdoctoral Associate, Genome Sciences, University of Washington
Dates: 08/05–05/08
Prior degree: Ph.D., Computer Science, Cornell University, 2005
Project: "Combining sequence and structure information to characterize yeast protein complexes."
Current position: Hedge fund.

Shobhit Gupta, Postdoctoral Associate, Genome Sciences, University of Washington
Dates: 01/06–06/07
Prior degree: Ph.D., Bioinformatics, Max-Planck Institute for Molecular Genetics, Berlin, Germany
Project: "Discovery and characterization of regulatory motifs in DNaseI hypersensitive sites."
Current position: MBA program, Cornell University

Tobias Mann, Ph.D. student, Genome Sciences, University of Washington
Dates: 06/03–06/07
Prior degree: M.S., Electrical Engineering, University of Washington, 2002
Project: "A thermodynamic approach to PCR primer design."
Current position: Director of Bioinformatics, Progenity

Darrin Lewis, Ph.D. student, Computer Science, Columbia University
Dates: 07/00–07/06.
Prior degree: M.S., Computer Science, Hofstra University, 1993
Project: "Combining kernels for classification."
Current position: Postdoctoral Associate, Michael Zhang lab, Cold Spring Harbor Laboratory.

Asa Ben-Hur, Postdoctoral Associate, Genome Sciences, University of Washington
Dates: 05/04–07/05
Prior degree: Ph.D., Information Systems, Technion, Israel, 2001
Project: "SVM prediction of protein-protein interactions."
Current position: Associate Professor, Department of Computer Science, Colorado State University, Ft.

Collins, CO.

Wei Wu, Postdoctoral Associate, Genome Sciences, University of Washington
Dates: 04/03–04/05
Prior degree: Ph.D., Biomedical Sciences, SUNY Albany, 2003, and M.S., Computer Sciences, SUNY Albany, 2002
Project: "Data fusion for cell cycle gene recognition in yeast."
Current position: Microsoft, Redmond, CA.

Paul Pavlidis, Associate Research Scientist, Columbia Genome Center
Dates: 12/99–6/02
Prior degree: Ph.D., Molecular and Cell biology, UC Berkeley, 1994, and five years of postdoctoral experience in neuroscience at Stanford (Daniel V. Madison) and Columbia (Eric R. Kandel)
Project: "Statistical and machine learning tools for microarray analysis."
Current position: Professor, Department of Psychiatry, University of British Columbia.

Victoria Haghighi, Postdoctoral Associate, Columbia Genome Center
Dates: 02/00–6/02
Prior degree: B.A., computer science, UC Berkeley, 1993, and Ph.D., genetics and development, Columbia, 2000.
Project: "Gene finding and genetic network modeling."
Current position: Assistant Professor, Department of Psychiatry, Columbia University.

Li Liao, Masters student, Computer Science, Columbia University
Prior degree: Ph.D., physics, Peking University, 2002. Dates: 9/01–6/02
Project: "Combining pairwise sequence similarity and support vector machines for remote protein homology detection."
Current position: Associate Professor, Department of Computer Science, University of Delaware.

Haoyuan Zhu, Masters student, Computer Science, Columbia University
Dates: 01/01–06/02
Project: "Software tools for the analysis of promoter regions."
Current position: Rosetta Inpharmatics, Kirkland, WA.

Johnny Li, Masters student, Computer Science, Columbia University
Dates: 09/01–06/02
Prior degree: Ph.D., Molecular Genetics, UT Austin, 1998
Project: "Statistics of replicated microarray experiments."
Current position: Scientist—Bioinformatics & Biostatistics, Nestlé Research Center, St. Louis, MO.

Ilana Hefter, Masters student, Computer Science, Columbia University
Dates: 09/01–06/02
Project: "Computational methods to differentiate spliced exons from pseudoexons."
Subsequent position: medical school.

Eleazar Eskin, Ph.D. student, Computer Science, Columbia University
Dates: 11/99–05/02
Project: "Sparse sequence modeling with applications to computational biology and intrusion detection."
Current position: Associate Professor, Department of Computer Science and Engineering, University of California, Los Angeles.

Jie Qin, Masters student, Computer Science, Columbia University
Dates: 06/00–06/01
Prior degree: Ph.D
Project: "Web-based microarray analysis tools and kernel hierarchical clustering for microarray analysis."
Current position: Associate Research Scientist, Columbia Genome Center.

James Cai, Masters student, Medical Informatics, Columbia University

Dates: 06/00–6/01
Prior degree: Ph.D., Biochemistry, University of Vermont
Project: "Cancer classification from microarray data."
Current position: Roche Pharmaceuticals.

Chihiro Ishii, Masters student, Computer Science, Columbia University
Dates: 3/00–12/00
Project: "Meta-MEME modeling of promoters and proteins."
Current position: NEC Tokyo.

Thesis committee membership

Adam Wilcox. *Automated classification of medical text reports.* Department of Medical Informatics, Columbia University, 2000.

Regina Barzilay. *Multidocument summarization by information fusion.* Department of Computer Science, Columbia University, 2001.

Heather McCune. *Better late than never: microarray-based analysis of DNA replication dynamics in a yeast mutant with inefficient late origin activation.* Department of Genetics, University of Washington, 2004.

Amol Prakash. *Algorithms for comparative sequence analysis and comparative proteomics.* Department of Computer Science and Engineering, University of Washington, 2006.

Nick Norberg. *Giant Zeeman effects in colloidal diluted magnetic semiconductor quantum dots with homogeneous dopant speciation.* Department of Chemistry, University of Washington, 2006.

Rui Kuang. *Inferring protein structure with discriminative learning and network diffusion.* Department of Computer Science, Columbia University, 2006.

Luca Cazzanti. *Generative models for similarity-based classification.* Department of Electrical Engineering, University of Washington, 2007.

Cristi Stoick-Cooper. *The role of Wnt signaling in fin and heart regeneration of the zebrafish.* Program in Neurobiology and Behaviour, University of Washington, 2007.

Chris Bartels. *Graphical models for large vocabulary speech recognition.* Department of Electrical Engineering, University of Washington, 2008.

Paul Murphy. Program in Molecular and Cellular Biology, University of Washington, 2008.

Chua Hon Nian. *Graph-based Methods for Protein Function Prediction.* Graduate School for Integrative Sciences and Engineering, National University of Singapore, 2008.

Zizhen Yao. *Genome scale search of noncoding RNAs: bacteria to vertebrates.* Department of Computer Science and Engineering. University of Washington, 2008.

Michael Hoopman. *Identification of unusual isotope distributions to facilitate rapid, high-throughput analysis using mass spectrometry.* Department of Genome Sciences, University of Washington, 2009.

Kevin Duh. *Semi-supervised ranking: algorithms & applications.* Department of Electrical Engineering, University of Washington, 2009.

Will Sheffler. *A volumetric score function for computational protein structure modeling and protein structure validation.* Department of Genome Sciences, University of Washington, 2009.

Eithon Cadag. *Statistical learning and data integration across heterogeneous sources for virulence factor recognition.* Department of Biomedical and Health Informatics, University of Washington, 2009.

Jon Malkin. *Smoothing class transitions with hard labels.* Department of Electrical Engineering, University of Washington, 2009.

Youyi Fong. *Modeling protein subfamilies; finding the number of mixture components from sequences of*

generalized Bernoulli random variables. Department of Biostatistics, University of Washington, 2009.

Yihua Chen. *Strategies for Similarity-based Learning.* Department of Electrical Engineering, University of Washington, 2010.

Richard Schaeffer. *Detecting conservation of folding pathway by family: Three-helix bundles, SH3 domains and β-graph domains.* Department of Biochemistry, University of Washington, 2010.

Alex Scouras. Department of Biochemistry, University of Washington. 2010.

Sangsoon Woo. *A computationally efficient modular optimal discovery procedure.* Department of Biostatistics, University of Washington, 2010.

Troy Zerr. *Genotyping Human Genomic Structural Variation.* Department of Genome Sciences, University of Washington, 2010.

Pragya Singh. *Structural and Functional Characterization of Bacterial Secretion Systems Using Mass Spectrometry.* Department of Medicinal Chemistry, University of Washington, 2011.

James Thompson. *Comparative Modeling of Protein Structures.* Department of Genome Sciences, University of Washington, 2011.

Claire Ryu. *Statistical Approaches to Analyze Mass Spectrometry Data.* Department of Statistics, University of Washington, 2011.

Albert Huang. *Integrated systems approach to the study of dendritic cell differentiation.* Program in Molecular and Cellular Biology, University of Washington, 2011.

Jarrett Egertson. *Development of Data Independent Acquisition Techniques for the Analysis of High-Complexity Samples by Tandem Mass Spectrometry.* Department of Genome Sciences, University of Washington, 2013.

Sharon Greenblum. *Metagenomic systems biology: frameworks for modeling and characterizing the gut microbiome.* Department of Genome Sciences, University of Washington, 2014.

Jeff Staples. *PRIMUS: Pedigree Reconstruction and Identification of a Maximum Unrelated Set.* Department of Genome Sciences, University of Washington, 2014.

Marea Cobb (masters). *NGSdb: A NGS Data Management and Analysis Platform for Comparative Genomics.* Department of Biomedical and Health Informatics, University of Washington, 2015.

Alexander Rosenberg. *Learning Models of Gene Expression from Synthetic DNA Sequences.* Department of Electrical Engineering, University of Washington, 2016.

Kai Wei. *Submodular Optimization and Data Processing.* Department of Electrical Engineering, University of Washington, 2016.

Ying Sonia Ting. *Shifting the Paradigm: Peptide-Centric Analysis of Systematically Sampled Mass Spectrometry Data.* Department of Genome Sciences, University of Washington, 2016.

Jean Morrison. *Statistical Methods for Spatially Structured Genomic Data and High Dimensional Studies.* Department of Biostatistics, University of Washington, 2016.

Sergey Ovchinnikov. *Protein Structure Determination Using Metagenomic Sequencing Data.* Program in Molecular and Cellular Biology, University of Washington, 2017.

Rodrigo Goya. *Bioinformatics Approaches for Identifying Single Nucleotide Variants and Profiling Alternative Expression in Cancer Transcriptomes.* Program in Bioinformatics, University of British Columbia, 2017.

Sen Zhao. Department of Biostatistics, University of Washington.

Han-Yin Yang. *Development Of Amyloidosis Typing Method and Data Acquisition Strategy Using Tandem Mass Spectrometry.* Department of Genome Sciences, University of Washington, 2018.

Brian Searle. *Development Of Data Independent Acquisition Methods to Systematically Analyze the human proteome.* Department of Genome Sciences, University of Washington, 2018.

Xioajie Qiu. *A Computational Framework to Resolve Complex Developmental Trajectories and Regulatory Networks from scRNA-seq.* Department of Genome Sciences, University of Washington, 2018.

Sumit Mukherjee. Department of Electrical Engineering, University of Washington. 2018.

Seungsoo Kim. *Maps And Mechanisms Of Three-dimensional Genome Organization.* Department of Genome Sciences, University of Washington, 2019.

Cecilia Noecker. *Metabolic Modeling-based Tools For Integrative Microbiome Data Analysis.* Department of Genome Sciences, University of Washington, 2019.

Alex Eng. *Model-based Computational Methods to Aid the Design of Synthetic Microbial Communities.* Department of Genome Sciences, University of Washington, 2019.

Hannah Pliner. *Algorithms for Modeling Gene Regulation and Determining Cell Type Using Single-cell Molecular Profiles.* Department of Genome Sciences, University of Washington, 2019.

Natalie Sauerwald. *Algorithms for the study of chromosomal structure variability.* CMU-Pitt Computational Biology Program, Carnegie Mellon University. 2020.

Ken Jean-Baptiste. *Root Development in* Arabidopsis thaliana. Department of Genome Sciences, University of Washington. 2020.

Daniel Jones. *RNA-seq analysis using approximate likelihood.* Paul G. Allen School of Computer Science and Engineering, University of Washington. 2020.

Eliah Overbey. *Understanding Regulatory Control of Gene Expression using Epigenome and Epitranscriptome Maps.* Department of Genome Sciences, University of Washington, 2021.

Mitchell Vollger. *Assembly of Segmental Duplications and Their Variation in Humans.* Department of Genome Sciences, University of Washington. 2021.

Jared Mohr. *New Technologies for Cross-linking Mass Spectrometry.* Department of Genome Sciences, University of Washington, 2022.

Betul Akgol Oksuz. *Detection of 3D Genome Folding at Multiple Scales.* University of Massachusetts Medical School, 2022.

Anthony Barente. *Powering Phosphoproteomics with Large Scale Data Analysis and Machine Learning.* Department of Genome Sciences, University of Washington, 2022.

Michael Goldberg. *Variation in Germline Mutagenesis in Humans and Other Great Apes.* Department of Genome Sciences, University of Washington, 2022.

Steven Wu. *Epigenomic Profiling of Human Tissues and Single-cell Resolution.* Molecular Engineering Pogram, University of Washington, 2022.

Lilian Heil. *Systematic Data Acquisition and Analysis Strategies for Quantitative Proteomics.* Department of Genome Sciences, University of Washington, 2023.

Kathleen Abadie. *Control of Stemness in CD8 T Cell Differentiation.* Department of Bioengineering, University of Washington, 2023.

Deanna Plubell. Department of Genome Sciences, University of Washington.

Hyeon-Jin Kim. Department of Genome Sciences, University of Washington.

Chengxiang Qiu. Department of Genome Sciences, University of Washington.

Sean Windle. Department of Pathology, University of Washington.

Miranda Mudge, Department of Genome Sciences, University of Washington.

Morgan Hamm, Department of Genome Sciences, University of Washington.

Tony Li, Department of Genome Sciences, University of Washington.

Mohsen Ferdosie, Department of Computational Biology, Carnegie Mellon University.

Abbey Thorpe, Department of Genome Sciences, University of Washington.

Mohamed Adil, Department of Lab Medicine Pathology, University of Washington.

Malvika Tejura, Department of Genome Sciences, University of Washington.

Joe Min, Department of Genome Sciences, University of Washington.

Teaching

"GENOME 373: Genomic Informatics." This course provides a survey of bioinformatics methodologies. The course is taught in an alternating format, with one week focusing on the application of bioinformatics programs, followed by a week in which I describe the inner workings of those same programs. First offered spring quarter 2004, I taught the course jointly with Jim Thomas for four years, then we did not teach the course in 2008, and I taught it myself in 2009.

"GENOME 475: Debates in Genetics." This course is organized into three-class units, where each unit addresses a current topic on genetics. The introductory class involves a lecture and discussion of the topic background, with readings from the popular press and the scientific literature. The second class is a formal debate, with student teams debating and students evaluating the debate. The third class is a discussion of the debate. I co-taught the course for five years (2012–2016).

"GENOME 541: Introduction to Computational Molecular Biology: Genome and Protein Sequence Analysis." This is the second half of a two-quarter graduate course in the core methods of computational biology. The course, which I organize, covers a variety of topics, with each of five faculty members teaching a two-week block. My portion of the course focuses on machine learning methods for computational biology. I have taught this course every year since 2003, except during my sabbaticals in 2010 and 2017.

"GENOME 559: Introduction to Statistical and Computational Genomics." This was originally a 5-week course that introduced graduate students to the basics of statistical analysis and programming in Perl. I taught the course for seven years from autumn 2002 through spring 2009. In 2007, we switched from Perl to Python, and in 2008, we switched from a 5-week to a 10-week format and a co-teaching format. I started teaching the course again in winter 2018.

"CSE590C: Reading and Research in Computational Biology." This ongoing seminar is offered through the UW Department of Computer Science and serves as the journal club for the Computational Molecular Biology program. The course is organized jointly by Martin Tompa, Larry Ruzzo, Joe Felsenstein and myself. I have taught this course every year since 2002–2003, except during my sabbaticals spring 2010 and 2017.

"Introduction to Bioinformatics." Three-week course taught at the African Institute for Mathematical Sciences, Muizenberg, South Africa, in February, 2013. AIMS draws post-baccalaureate students from across Africa and provides them with a one-year masters degree to prepare them for entry to graduate school. My course taught basic bioinformatics using Python.

Ad hoc teaching at the University of Washington:

- "MEDCHEM 541: Mass Spectrometry Based Proteomics." 2008, 2009.
- "MEBI/BIME 590: Biomedical and Health Informatics Lecture Series." 2008, 2009, 2011, 2017.
- Data science masters program, 2016, 2017.

"Computational Genomics." I designed this course to be offered to undergraduate and graduate students at Columbia. It was cross-listed in computer science, biomedical engineering and medical informatics and

was offered as an elective in all three departments. The course focused on microarray expression analysis, protein sequence analysis, and gene finding. Students worked in interdisciplinary teams on two projects per semester. I taught the course in spring 2000 (35 students) and 2001 (65 students). After I left Columbia, the course was taught by Prof. Christina Leslie.

"Data Structures and Algorithms in Java." This is a core course in the undergraduate computer science department curriculum at Columbia. I taught the course in autumn 1999 (50 students), 2000 (60 students) and 2001 (71 students).

Participant, Preparing Professional Faculty Initiative. University of California, San Diego, winter, 1998. Prof. Clark Glymour, faculty mentor.

Teacher, math, physics and English literature. Mt. Tabor High School, Lesotho, 1991-1993.

Departmental service

Department of Genome Sciences

- Building committee, 2002–2003.
- Faculty search committee, 2004–2005, 2007–2008, 2017–2018, 2018–2019, 2022–2023, 2023-2024.
- Admissions committee, 2004–2006.
- Seminar committee, 2006, 2007, 2014, 2019.
- Curriculum committee, 2006, 2010–.
- Chair, faculty search committee, 2011–2012.
- Chair, promotions committee, 2018–2019.
- Co-organizer, departmental symposium, 2007, 2024.
- Chair, IT committee, 2018–.
- Chair, DEI committee, 2020–.
- Interim Chair, 2020–2021.
- Faculty mentor, Elhanan Borenstein, Su-In Lee, Doug Fowler, Devin Schweppe.

University of Washington

- Faculty search committee, Institute for Protein Design, 2015
- Biostatistics Program Review Committee, Autumn 2024

Training

- Nov 19, 2019: One-day mentorship training program run by the Center for the Improvement of Mentored Experiences in Research (`https://cimerproject.org`).
- February 8, 2020: Ally skills workshop run by Frame Shift Consulting (`https://frameshiftconsulting.com/`). 3 hours.
- May 26, 2020: "Moving from bystander to upstander: take action to combat harassment and aggression." NIH webinar. 2 hours.
- Sep 2, 2020: "Oh! So, You Really Want to Talk about Race? : Racial, Health and Environmental Justice in STEM." PRISM workshop, 1.5 hours.

- Sep 4, 2020: "From Deficits to Possibilities: Mentoring Lessons for Building A Culture of Broad Access and Success." PRISM workshop, 1.5 hours.

- Oct 20, Nov 10, 2020: "Race, Bias & Dissonance" run by Greg Taylor (`http://www.cconnectconsult.com`). Two 2-hour sessions.

- Oct 30, 2020: "Leadership Development Series: Working Through Conflict." UW SoM workshop. 2 hours.

- Feb 23, March 23, April 27, 2021: "Anti-racism training series," UW SoM. Three 2.5-hour sessions.

- Mar 22, 2021: "Leadership Development Series: Creating High Performing Teams." UW SoM. 2 hours.

- Apr 14, 2021: "Virtual Summit on Equity in Graduate Education." UW Graduate School. 1.5 hours

- May 18, 2021: "So You Want to Talk About Race." Ijeoma Oluo webinar. 1 hour.

- May 20, 2021: "Holistic Support of the Graduate Student Experience: Towards Effective Mentorship Practices." UW Office of Equity & Justics in Graduate Programs. 1.5 hours.

- Feb 16, Mar 9, June 15, 2021: "99.9% Alike: Making Genomics Inclusive for Marginalized Populations." Lavender Rights Project. Three 2-hour sessions.

- Mar 7–11, 2022: "Facilitating Entering Mentoring." Facilitator training program run by the Center for the Improvement of Mentored Experiences in Research. Six 2-hour sessions.

- Oct-June, 2022–2023: "Gilliam Mentorship Skills Development Training."

- May, 2023: "Entering Mentoring" workshop and discussion group. Co-facilitator for five 2-hour sessions for Genome Sciences faculty.

- Apr 11-12, 2024: "Mentorship Development Coures." Howard Hughes Medical Institute. 9 hours.
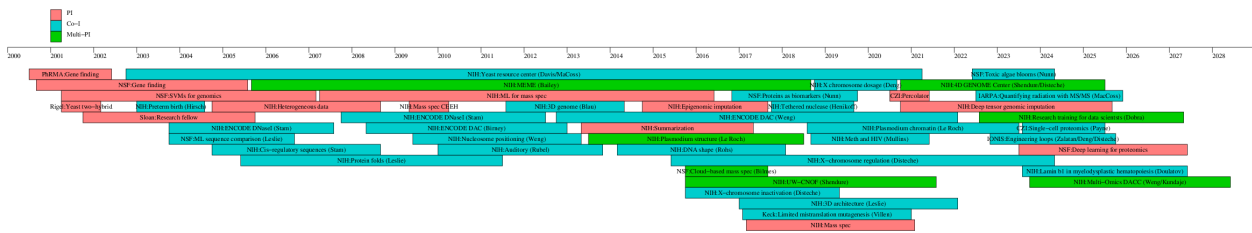
Current support

(Payne)	07/01/23–06/30/25
Chan-Zuckerberg Initiative	$66,667

Improved quantitative accuracy for single cell proteomics
The overall goal of this project is to enable accurate protein quantification from single-cell tandem mass spectrometry data by enhancing the existing match-between-runs approach, implementing the proposed algorithms in publicly available software for use by the broader research community.
Role: Co-I

R25 HG012337-01 (Dobra, Noble)	07/01/22–04/30/27
NIH/NHGRI	$1,260,275

University of Washington Genomics Research Training for Data Scientists
In support of the NHGRI Strategic Plan to Diversify the Genomics Workforce, we propose the University of Washington Genomics Training for Data Scientists program (UWGRTDS). This interdisciplinary program provides masters students in data science from underrepresented minority (URM) backgrounds with paid research experiences in genomics research laboratories at UW.
Role: Co-I

(Zalatan, Deng, Disteche)   10/01/22–9/31/25
IONIS Pharmaceuticals $35,932
Engineering loops and repositioning chromatin to express or silence disease-associated X-linked genes
Role: Co-I

W911NF-22-2-0059 (MacCoss)   06/01/22–11/30/25
IARPA $639,949
Biodosimetry: Quantitation of radiation induced protein adducts using mass spectrometry
The goals of the TEI-REX program are to discover, validate, and define a set of biomarkers that can be effectively used to build classifiers for biodosimetry.
Role: Co-I

R21 ES034337 (Nunn)   5/01/22–4/30/24
NIH $31,766
Using microbiomes as microsensors to forecast toxic algae blooms
This project will use metaproteomic and metabolomic methods to establish marine microbiomes as microsensors for use in an early detection system to warn the public before the onset of toxic algal blooms, thereby avoiding exposure.
Role: Co-PI

R01 HG011466 (Noble)   2/01/21–1/31/25
NIH $1,549,944
Deep tensor genomic imputation
This project will develop a machine learning approach to predicting various types of genomic experiments before they are performed by leveraging big, heterogeneous compendia of existing data.
Role: PI

UM1 HG011531 (Noble, Disteche, Shendure)   9/18/20–6/31/25
NIH $10,155,255
UW 4-Dimensional Genomic Organization of Mammalian Embryogenesis Center
The UW 4D Genome Organization of Mammalian Embryogenesis Center (UW 4D GENOME) will carry out systematic generation of sequencing and imaging data during mouse embryogenesis, summarizing and visualizing the resulting data using machine learning models.
Role: Multi-PI

R35 GM131745 (Disteche)   5/1/19–4/30/24
NIH/NIGMS $2,912,429
X chromosome regulation and role in aneuploidy
The main goal of this project is to understand the role of the sex chromosomes in sex differences and sex chromosome disorders in vivo. The Noble lab is developing statistical and machine learning methods to interpret the resulting data sets.
Role: Co-investigator

2245300   6/1/2023–5/31/27
NSF $1,199,760
DMS/NIGMS 2: Deep learning for repository-scale analysis of tandem mass spectrometry proteomics data
The primary technology driving the rapid growth of proteomics is tandem mass spectrometry. In addition to technological advances in mass spectrometry hardware, accurate and efficient analysis of the complex data produced by a tandem mass spectrometer requires increasingly sophisticated computational tools. The proposed project will develop these tools.
Role: PI

R01 HL169156 (Doulatov)   7/1/2023–5/31/27
NIH/NLHBI $615,675 (Noble lab)
The role of lamin B1 in normal and myelodysplastic hematopoiesis
Role: Co-investigator

U01 HG013198 (Weng)   9/1/2023–5/31/28
NIH/NHGRI $408,426 (Noble lab)
Multi-Omics DACC: The Data Analysis and Coordination Center for the collaborative multi-omics for health and disease initiative
The overall goal of the Multi-Omics for Health and Disease Consortium is to explore the use of multiomics data to detect and assess molecular profiles associated with healthy and disease states, coalescing these data into a multi-dimensional dataset that is available to the research community
Role: Co-investigator

Previous support

R01 AI136511 (Le Roch)   7/01/18–4/30/23
NIH/NIGMS $764,953 (subcontract)
Chromatin structure and control of gene expression in the human malaria parasite
This project will characterize the molecular drivers of chromatin organization and gene regulation in *Plasmodium falciparum*, elucidating their importance in parasite development, virulence, and sexual differentiation.
Role: Co-investigator

2020-219003 (Noble)   7/01/20–6/30/22
Chan-Zuckberberg $200,000
The Percolator analysis engine for tandem mass spectrometry data
This project improves Percolator, which is the dominant software for analyzing spectrum identifications produced by protein tandem mass spectrometry, by making the software faster, more robust, and applicable to more types of mass spectrometry data. A total of 23 out of 194 (11.9%) of proposals were funded.
Role: PI

P41 GM103533 (MacCoss)   7/1/17–3/31/21
NIH/NIGMS $856,289 (subaward)
Comprehensive biology: Exploiting the yeast genome
The mission of the YRC is to facilitate the identification and characterization of protein complexes in the yeast *Saccharomyces cerevisiae.*
Role: Co-investigator

R01 GM121818 (Noble)   2/01/17–1/31/21
NIH/NIGMS $1,279,410
Optimization and joint modeling for peptide detection by tandem mass spectrometry
This project develops and applies statistical and machine learning methods for interpreting mass spectra.
Role: PI

U01 HG009395 (Leslie)   12/1/16–1/31/22
NIH/NHGRI $380,260 (subcontract)
Encoding genomic architecture in the encyclopedia: linking DNA elements, chromatin state, and gene expression in 3D
This project developed methods to integrate gene expression, DNA sequence, chromatin accessibility, and chromatin 3D architecture data to gain insights into gene regulation.
Role: Co-investigator

U24 HG009446 (Weng)   2/1/17–1/31/22
NIH/NHGRI $550,000 (subcontract)
EDAC: ENCODE Data Analysis Center
The EDAC is a multi-disciplinary group of scientists who integrate data from all groups in the
ENCODE Consortium.
Role: Co-investigator

R61 DA047010 (Mullins, Duan)   8/15/18–5/31/22
NIH/NIGMS $204,939 (subaward)
**Impact of methamphetamine use on the HIV nucleome in individuals on antiretroviral
therapy**
This project aims to understand the mechanisms of HIV latency using state-of-the-art genomic tools
as well as complementary cellular and molecular methods.
Role: Co-investigator

R01 GM127327 (Deng)   9/01/18–8/31/21
NIH $85,335 (subaward)
Gene-by-gene studies of dosage regulation pathways of the mammalian X chromosome.
The goal of this study is to identify mechanisms of adaptation in response to gene dosage changes
such as copy number variants, using the X chromosome as a model.
Role: Co-investigator

Grant (Villen)   1/1/17–6/30/21
Keck Foundation $185,745 (subaward)
**Limited mistranslation mutagenesis: A proteome-wide technology to assess the effects
of amino acid substitutions on protein function**
This project will develop a novel, non-genetic mutagenesis technology to assess the effects of amino
acid substitutions on protein function on a proteome-wide scale
Role: Co-investigator

U54 DK107979 (Noble, Shendure)   9/30/15–7/31/21 $11,444,752
NIH
University of Washington Center for Nuclear Organization and Function
This project developed novel experimental and computational methods to characterize genome 3D
architecture, validate the methods using mouse and human cells, and demonstrate the utility of the
resulting data for improving our understanding of fundamental biology and human disease.
Role: Multi-PI

1633939 (Nunn)   10/1/16–9/30/20 $410,284 (subaward)
NSF
**Collaborative research: Proteins as functional biomarkers: integrating organic char-
acterization with proteomics to track routes for carbon and nitrogen recycling and
preservation**
This project aims to produce a metaproteomic pipeline to provide insights into the distribution of
taxonomy-specific biological functions to unravel ecosystem processes.
Role: co-PI

T32 CA206089 (Noble, Daniel, Fairhall, Witten)   9/23/15–8/31/20 $1,343,450
NIH
University of Washington PhD Training in Big Data for Genomics and Neuroscience
This training grant provides predoctoral training for students engaged in big data research in ge-
nomics and neuroscience at the University of Washington.
Role: Multi-PI

U01 CA200147 (Henikoff, Shendure, Noble) 9/1/17–8/31/19 $101,629
NIH
Tethered nuclease strategies for *in situ* mapping of 3D nuclear organization
This proposal refines and extends a modern genomic profiling method, called CUT&RUN, that
enables accurate mapping with relatively low-coverage deep sequencing.
Role: Co-investigator

R01 RR021692, R01 GM103544 (Noble, Bailey) 8/01/05–8/31/18 $2,859,555
NIH/NIGMS
The MEME suite of motif-based sequence analysis tools
This project supports, maintains and develops the MEME software suite of motif analysis software.
Role: Multi-PI

R01 AI106775 (Noble, Le Roch) 6/7/13–5/31/18 $1,756,786
NIH/NHGRI
The spatial organization of the *Plasmodium* genome throughout its infectious cycle
This project, a collaboration with Karine Le Roch's lab at UC Riverside, used a recently developed
technique to discover the three-dimensional structure of the *P. falciparum* genome at three time
points during the parasite's life cycle, assay the chromatin modification landscape of the genome
over the same time points, and then built an integrated model of gene expression.
Role: Multi-PI

R01 GM106056 (Rohs) 2/1/14–1/31/18 $157,590 (subcontract)
NIH
Genome analysis based on the integration of DNA sequence and shape
This project carries out integrated analysis of 1D sequence at genome-wide scale with 3D structure
at atomic resolution, developing novel genome analysis tools and furthering our understanding of
genome function.
Role: Co-investigator

U41 HG007000 (Weng) 9/12/12–7/31/17 $721,688 (subaward)
NIH/NHGRI
EDAC: ENCODE Data Analysis Center
This proposal aims to create the ENCODE Data Analysis Center consisting of a multi-disciplinary
group of leading scientists who will respond to directions from the ENCODE Analysis Working
Group and thus integrate data from all groups in the ENCODE Consortium in an unbiased manner.
The Noble lab developed unsupervised and semi-supervised learning methods for interpreting the
data tracks.
Role: Co-investigator

1549932 (Noble, Bilmes, MacCoss) 9/1/15–8/31/17 $61,534
NSF
EAGER: Cloud-based analysis of mass spectrometry proteomics data
This project will develop novel machine learning algorithms to enable proteomics researchers to
make sense of one experiment within the context of all the previous experiments stored in the cloud.
Role: Multi-PI

R01 ES024917 (Noble) 9/10/14–8/31/17 $94,139
NIH/NIEHS
Machine learning methods to impute and annotate epigenomic maps
This project develops and applies novel unsupervised machine learning methods to make sense of
epigenomic data sets. The project will produce novel computational methods as well as two primary
analysis products: virtual experiments for combinations of assays and cell types that have not yet
been carried out and annotations that identify various types of biochemical and functional activity
along the human genome.
Role: PI

R01 CA180777 (Noble)   4/1/13–3/31/17 $76,247
NIH/NCI
BIGDATA: Small: DA: Interpreting massive genomic data sets via summarization
This proposal develops submodular summarization methods for drawing inferences from massive
genomic data sets. We applies these methods to five big data problems in genomics: 1) identifying functional elements characteristic of a given human cell type; 2) identifying genomic features
associated with particular subclasses of cancer; 3-4) identifying genomic variants representative of
ancestrally or phenotypically defined human populations; and 5) finding a set of microbial genes
that characterize a given site on the human body.
Role: PI

R01 GM096306 (Noble)   6/01/12–5/31/16 $2,483,909
NIH/NIGMS
Machine learning analysis of tandem mass spectra
This project applies techniques and tools from the field of machine learning to the analysis of mass
spectrometry data. The aims include developing better methods for identifying and quantifying
proteins in complex mixtures, including proteins with post-translational modifications.
Role: PI

R01 GM098039 (Blau)   7/01/11–4/30/14 $1,801,829
NIH/NIGMS
Modeling mammalian genomes
In this project we used our previously described genome architecture assay to understand the relationship between genome architecture and allele specific expression in mammalian development.
Role: Co-investigator

R01 DC03829 (Rubel)   12/1/09–10/31/13 $9520 (salary only)
NIH/NIDCD
Afferent influences on auditory system ontogeny
The overall goals of this program are to understand the cellular basis of experience-dependent maturation and maintenance of neurons in the brainstem auditory pathways of birds and mammals.
Role: Co-investigator

DBI 085008 (Weng)   5/15/09–4/31/13 $292,803
NSF
Computational analysis of nucleosome positioning signals in eukaryotic genomes
This project aims to develop machine learning methods that model the relationship between DNA
sequence and local chromatin structure.
Role: Co-investigator

U54 HG004695 (Birney)   4/1/08–12/31/12 $630,978
NIH/NHGRI
EDAC: ENCODE Data Analysis Center
This project established a multi-disciplinary group of scientists to provide integrative analysis of
functional information from the NIH ENCODE consortium.
Role: Co-investigator

U65 HG004592 (J Stamatoyannopoulos)   9/29/07–6/30/12 $13,971,444
NIH/NHGRI
A comprehensive catalog of human DNaseI hypersensitive sites
The overall aim of this proposal is to establish a comprehensive, high quality catalog of human
DNaseI hypersensitive sites spanning all major tissue lineages.
Role: Co-investigator

P41 RR0011823 (Davis)   9/1/06–8/31/11 $2,427,687
NIH/NCRR
Comprehensive biology: Exploiting the yeast genome
The mission of the YRC is to facilitate the identification and characterization of protein complexes
in the yeast *Saccharomyces cerevisiae*.
Role: Co-investigator

R01 GM074257 (Leslie)   5/1/05–6/30/11 $211,974 (subcontract)
NIH/NCRR
Recognizing protein folds with discriminative learning
This project develops discriminative methods for classifying proteins into structural families based
upon their amino acid sequences.
Role: Co-investigator

R01 EB007057 (Noble)   3/1/07–2/28/11 $2,427,687
NIH/NIBIB
Machine learning analysis of tandem mass spectra
This project applies techniques and tools from the field of machine learning to the analysis of
mass spectrometry data. The primary aim is to produce software that increases the sensitivity and
specificity of protein identifications from complex mixtures.
Role: PI

Grant (Noble)   4/1/09–3/31/10 $24,816
NIH
**Facilitating and improving mass spectometry analyses in the Center for Ecogenetics
and Environmental Health**
This project improved the documentation and user interface for the Crux mass spectrometry analysis
toolkit and applied Crux to several data sets being produced under the auspices of the CEEH.
Role: PI

R33 HG003070 (Noble)   9/1/04–8/31/08 $1,226,036
NIH/NHGRI
Detecting relations among heterogeneous genomic datasets
This project developed a computational framework for integrating and drawing inferences from a
collection of genome-wide measurements. The project included subcontracts at three other institu-
tions.
Role: PI

R01 GM071923 (J Stamatoyannopoulos)   9/1/04–8/31/08 $1,746,932
NIH/NIGMS
Computational discovery of *cis*-regulatory sequences
This project used quantitative chromatin profiling to identify *cis*-regulatory elements in a high-
throughput fashion.
Role: Co-investigator

U01 HG003161 (G Stamatoyannopoulos)   9/30/03-7/31/07
NIH/NHGRI
Identification of Functional DNA Elements by HSqPCR
This was an ENCODE project, aimed at detecting DNaseI hypersensitive sites *in vivo* using a high-
throughput screen.
Role: Co-investigator

IIS-0093302 / IIS-0431725 (Noble)   3/1/01-2/28/07 $445,074 (direct)
NSF
CAREER: Support vector methods for functional genomics
This project develops gene functional classification techniques based upon the support vector machine algorithm.
Role: PI

P41 RR0011823 (Davis)   9/1/01–8/31/06
NIH/NCRR
Comprehensive biology: Exploiting the yeast genome
The mission of the YRC is to facilitate the identification and characterization of protein complexes in the yeast *Saccharomyces cerevisiae.*
Role: Co-investigator

EIA-0312706 (Leslie)   9/15/03-8/31/06 $200,000 (direct)
NSF
ITR: Machine learning approaches to protein sequence comparison: discriminative, semi-supervised, scalable algorithms
This project developed, implemented and disseminated machine learning algorithms for pairwise biological sequence comparison.
Role: Co-investigator

BR-4031T (Noble)   9/16/01-9/15/05 $40,000
Sloan
Alfred P. Sloan Foundation Research Fellow
The Sloan Research Fellowships were established in 1955 to provide support and recognition to young scientists. Currently, 100 grants are awarded annually in six fields: chemistry, computer science, economics, mathematics, neuroscience, and physics. In 2001, 14 grants were awarded in computer science.
Role: PI

BDI-0078523 / BDI-0243257 (Noble)   8/1/00-7/31/05 $412,195 (direct)
NSF
Generative and discriminative methods for gene finding and functional annotation
This project developed an integrated gene finding and protein homology detection system based upon a generalized expectation-maximization learning algorithm.
Role: PI

Grant (Grundy)   6/1/00-5/31/02 $60,000
Pharmaceutical Research and Manufacters of America Foundation
Data-driven computational gene finding and functional annotation
This award provided salary support for the development of gene finding and functional annotation techniques.
Role: PI

Grant (Grundy)   3/01/01-2/28/02 $91,580
Rigel, Inc.
Inferring gene function from yeast two-hybrid data
This project used yeast two-hybrid data, in conjunction with sequence and expression data, to functionally characterize yeast genes.
Role: PI

Grant (Honig)   6/1/01-5/31/04 $2,319,096
NIH
Center for Computational Biology and Bioinformatics
The Center for Computational Biology and Bioinformatics serves as a focal point for research and education at Columbia. The center provides core resources, sponsors retreats, workshops and a seminar series, and assists in the development of a coherent curriculum throughout the university.
Role: Co-investigator

BES-0088001 (Leonard)   1/1/01-12/31/03 $500,000 (direct)
NSF
Curriculum in Genomic Engineering
This project developed and operated a course track in Genomic Engineering that is related to research activities undertaken by the proposers and their colleagues at Columbia University. Prof. Noble taught "Computational Genomics," one of the five courses in this track, until leaving Columbia.
Role: Co-investigator

RO1 HD41689 (Hirsch)   12/1/02-7/31/04 $80,734 (subcontract)
NIH/NICHHD
The molecular pathenogenesis of health disparities in infection-induced preterm birth
This project used microarrays to identify genes involved in infection-induced preterm labor.
Role: Co-investigator