

Mark Gerstein

Williams Professor of Biomedical Informatics, Yale
Full CV as of 31 Dec. 2019 (with some sections less current)

Education

Harvard College	AB	1989	Physics (& History of Science)
Cambridge University	PhD	1993	Biophysics/Chemistry
Stanford University	post-doc	1993-1996	Bioinformatics

Positions

2006 -	AL Williams Prof. Biomedical Informatics, Yale U.
2002 -	co-director, Yale Computational Biology & Bioinformatics Program
2017 -	co-director, Yale Center for Biomedical Data Science
2006 -	Prof. Molecular Biophysics & Biochemistry, Yale U.
2006 -	Prof. Computer Science, Yale U.
2018 -	Prof. Statistics & Data Science, Yale U.
2001 - 2006	Assoc. Prof. Molecular Biophysics & Biochemistry and Computer Science, Yale U
1997 - 2001	Asst. Prof. Molecular Biophysics & Biochemistry, Yale U.

Honors

2015	ISCB (Intl. Society of Computational Biology) Fellow
2009	AAAS Fellow
1997 - 2001	Young Investigator Awards from Navy & IBM, and PhRMA, Donaghue, & Keck foundations
1993 - 1996	Damon Runyon-Walter Winchell post-doctoral Fellowship
1989 - 1993	Herchel-Smith Scholarship funded PhD at Cambridge
1989	Graduated college <i>summa cum laude & phi beta kappa</i>

Editorial Boards

Genome Research, Molecular Systems Biology, PLoS Comp Bio, GenomeBiology,
BMC Bioinformatics, Molecular & Cellular Proteomics, Protein Science,
Molecular Biology & Evolution, F1000 (co-head Big Data & Analytics Section)

Professional Experience (beyond Yale, but not including “for profits”)

Analysis Working Group co-chair: NHGRI ModENCODE Project ('07-'14), Brainspan Project ('09-'14),
1000 Genomes Functional Interpretation Group ('11-'15), exRNA consortium ('13-), ENCODE ('17-)
CMG [Centers for Mendelian Genomics] ('13-'18), PsychENCODE ('14-),
PCAWG-2 [Pan-Cancer Analysis Working Group, non-coding drivers] ('14-'19)

Committees: AnVIL External Consultant Committee ('19-), NE Big Data Innovation Hub (NSF-sponsored),
Governance Committee ('15-); Program Committee BIBM '09, '12, '15; NIH Human Proteome Meeting
Organizing Committee; International Pangolin Research Consortium (IPaRC) Advisory Committee;
Organizing Committee for NSF Workshop on Knowledge Management and Visualization Tools, '08

Teaching (as of 31 Dec. 2019)

Biomedical Data Science: Mining & Modeling

CBB752b, MBB752b, CPSC752b, MBB452, MBB753, MBB754, S&DS352
(preceding are Yale course numbers for Computational Biology (CBB), Biophysics (MBB),
CS (CPSC) and Statistics & Data Science (S&DS))

Principal instructor responsible for whole-semester course on fundamentals of bioinformatics
and biomedical data science taught to advanced undergraduates and graduate students.

Course comprised of 25 lectures of 75',
each with weekly section, graded homework and quizzes, midterm and final project.

Taught continuously for 23 iterations (since '98 through '20), usually in Spring.
(Name has been changed from Genomics & Bioinformatics, to Bioinformatics to its current name.)
Course web site is GersteinLab.org/courses/452

Other Miscellaneous Current Teaching

Integrated Workshop (on physics, engineering & biology topics)
(Yale course numbers: MBB 591a / ENAS 991a / MCDB 591a / PHYS 991a)
in fall '18, '19; responsible for 3 classes of 75'

Topics in Deep Learning: Methods & Biomedical Applications
(Yale course numbers: S&DS 567, MBB 567, CBB 567)
in spring '20; responsible for ~6 effective classes of ~75'

Miscellaneous Teaching in Past

Responsible Conduct of Research (MBB676b) in the Spring '14 & '15 (1 class)
Evidence Behind Health News (HLTH081) in Fall '16 (1 class)
Science and Politics of Cancer (MCDB 40) in Spring '17 (1 class)

More than 10 years ago taught parts of (eg ~6 75' lectures) :

- 1) CS Course "Introduction to Data Mining"
- 2) Molecular Biophysics course "Macromolecules"

Gerstein Lab Personnel (updated 12-Sept-2019)

PI(1)

Mark Gerstein	Albert L Williams Professor	01/1997
---------------	-----------------------------	---------

Laboratory Staff (2)

Mihali Felipe	Systems Administrator	11/2004
Lori Ianicelli	Administrative Assistant	03/2013

Research Scientists (7)

Joel Rozowsky	Research Scientist	09/2003
Sushant Kumar	Assoc. Res. Sci.	12/2013
Jing Zhang	Assoc. Res. Sci.	02/2014
Fabio Navarro	Assoc. Res. Sci.	10/2014
Shaoke Lou	Assoc. Res. Sci.	05/2015
Jonathan Warrell	Assoc. Res. Sci.	09/2016
Prashant Emani	Assoc. Res. Sci.	02/2017

Postdoctoral Associates and Fellows (9)

Leonidas Salichos		09/2014
Timur Galeev		09/2014
Jinrui Xu		11/2015
Gamze Gursoy		11/2016
Yucheng T. Yang		01/2018
Zhen Tan		01/2019
Garrett Ash	Associated Health Fellow	05/2019
Christopher J.F. Cameron	jt. w. Yong Xiong	06/2019
Kun Xiong		07/2019

Graduate Students (9)

Declan Clarke	Chemistry; short postdoc	08/2010
Michael R. Schoenberg	MBB (jt. w. M Simon); short postdoc	08/2012
Paul Muir	MCDB (jt. w. F Isaacs); short postdoc	08/2013
Donghoon Lee	CBB	08/2014
William Meyerson	CBB (MD/PhD)	08/2014
Xiaotong Li	CBB (jt. w. L Pusztai)	09/2014
Hussein Mohsen	CBB	08/2016
Jiahao Gao	CBB	09/2017
Tianxiao Li	CBB (CS master)	09/2017

Full-time Postgrads (2)

Jason Liu	full-time postgrad	05/2014
Charlotte Brannon	full-time postgrad	07/2019

Undergrads (9)

Zhanlin (Flynn) Chen		02/2018
George Gayed		06/2018
Mia Jackson		08/2018
Daniel Kim		11/2018
Andrew Zhang		02/2019
Justin Du		02/2019
John Leisman		05/2019
Edward Cho		07/2019
Yuhang Chen		08/2019

Misc (4)

Brian Barron	MD student	10/2015
Andrew Loza	Medical Resident	04/2018
Shua Kim	High school student	07/2018
Akshina Jindal	MS student intern	05/2019

Past Postdoctoral Associates and Fellows (updates on 12-Sept-2019)

Currently Holding a Faculty Position (26)

Hedi Hegyi	1998 - 2000	CEITEC	Group Leader
Jiang Qian	1999 - 2002	Johns Hopkins	Professor
Yuval Kluger	1999 - 2002	Yale University	Professor
Paul Harrison	1999 - 2004	McGill University	Assoc. Prof.
Nicholas Luscombe	2000 - 2004	UCL/The Francis Crick Institute	Professor
Zhaolei Zhang	2002 - 2004	University of Toronto	Professor
Alberto Paccanaro	2003 - 2005	CS Dept. Royal Holloway, U of London	Professor
Olof Emanuelsson	2003 - 2005	KTH Royal Institute of Technology, Sweden	Assoc. Prof.
Long (Jason) Lu	2003 - 2006	Cincinnati Children's Hospital	Assoc. Prof.
Yu (Brandon) Xia	2003 - 2006	McGill University	Professor
Deyou Zheng	2003 - 2007	Albert Einstein College of Medicine	Professor
Philip Kim	2004 - 2008	University of Toronto	Professor
Jan Korbel	2005 - 2007	EMBL	Group Leader
Zhengdong Zhang	2005 - 2010	Albert Einstein College of Medicine	Assoc. Prof.
Andrea Sboner	2006 - 2011	Weill Cornell Medicine	Asst. Prof.
Gang Fang	2007 - 2014	NYU Shanghai	Asst. Prof.
Zhi (John) Lu	2008 - 2011	Tsinghua University	Assoc. Prof.
Chao Cheng	2008 - 2012	Baylor College of Medicine	Assoc. Prof.
Alexej Abyzov	2008 - 2014	Mayo Clinic	Asst. Prof.
Ekta Khurana	2008 - 2014	Weill Cornell Medicine	Asst. Prof.
Arif Harmanci	2010 - 2017	UTHealth	Asst. Prof.
Cristina Sisu	2011 - 2017	Brunel University London	Asst. Prof.
Yan Zhang	2012 - 2016	Ohio State University	Asst. Prof.
Daifeng Wang	2012 - 2016	University of Wisconsin-Madison	Asst. Prof.
Daniel Spakowicz	2014 - 2018	The Ohio State University	Asst. Prof.
Chengfei Yan	2016 - 2019	Huazhong University of Science and Technology	Assoc. Prof.

Working in Industry (29)

Valery Trifonov	1998 - 2004	Goldman Sachs
Suganthi Balasubramanian	1999 - 2015	Regeneron Pharmaceuticals
Jochen Junker	2000 - 2002	Oswaldo Cruz Foundation
Ning Lan	2000 - 2002	GetHired Inc
Yang Liu	2000 - 2003	Sigma Aldrich
Ian Laurenzi	2002 - 2004	ExxonMobil
Sambath Chung	2002 - 2004	Genelogic
Ursula Lehnert	2002 - 2004	McKinsey Consulting
Duncan Milburn	2002 - 2005	Business Aspect
Chern-Sing Goh	2002 - 2006	Bioinformatics Consulting

John Karro	2003 - 2005	Google
Zhiyun (Eric) Yu	2003 - 2006	Matrix Partners China
Alexander Karpikov	2004 - 2007	Credit Suisse
Rajkumar (Raj) Sasidharan	2004 - 2008	Solvuu, Inc
Yongpan (Daniel) Yan	2005 - 2006	Glaxosmithkline
Can (John) Bruce	2005 - 2007	Sciomix
Thayalini Arinaminpathy	2005 - 2007	Detica Consulting
Anne (Counterman) Burba	2005 - 2009	Freelance consultant
Ashish Agarwal	2006 - 2010	Solvuu
Nitin Bhardwaj	2007 - 2011	Celmatix
Lukas Habegger	2007 - 2012	Regeneron Genetics Center
Roger Alexander	2007 - 2013	Pacific Northwest Diabetes Research Institute
Baikang Pei	2010 - 2016	Amgen Inc
Renqiang (Martin) Min	2011 - 2012	NEC Laboratories America
Robert Kitchen	2011 - 2016	Exosome Diagnostics
Wyatt Clark	2013 - 2014	BioMarin Pharmaceutical Inc
Anurag Sethi	2013 - 2016	Calico Life Sciences
Min Xu	2017 - 2018	XtalPi Inc.
Xiangmeng (Nelly) Kong	2017 - 2018	GIC
Xu Shi	2017 - 2019	AbbVie

Others (3)

Koon-Kiu Yan	2008 - 2017	St. Jude Children's Research Hospital
Shuang Liu	2015 - 2019	University of Wisconsin-Madison
Bian Li	2018 - 2019	Vanderbilt University

Past PhD students (updates on 12-Sept-2019)

Currently Holding a Faculty Position (5)

Paul Bertone	1998 - 2005	EBI (Cambridge)	Group Leader
Dov Greenbaum	1999 - 2004	IDC Herzliya	Professor
Haiyuan Yu	2000 - 2005	Cornell University	Professor
Samuel Flores	2004 - 2007	Stockholm University	Assoc. Prof.
Kevin (Yuk-Lap) Yip	2004 - 2009	The Chinese University of Hong Kong	Assoc. Prof.

Postdoc (2)

Lucas Lochovsky	2009 – 2018	The Jackson Laboratory	Postdoc
Shantao Li	2012 – 2018	Stanford U.	Postdoc

Working in Industry (19)

Werner Krebs	1996 - 2001	Acculation Inc	
Ronald Jansen	1997 - 2002	UBS Securities LLC	
Vadim Alexandrov	1998 - 2003	Psychogenics	
Rajdeep Das	1998 - 2004	WorldQuant	
Michael Seringhaus	2001 - 2007	Latham & Watkins LLP	
Andrew Smith	2002 - 2007	Bristol-Myers Squibb	
Thomas Royce	2002 - 2007	Illumina	
Yuen-Jong Liu	2004 - 2009	Sigma	
Jiang Du	2004 - 2010	BlueMountain Capital Management	
Hugo (Yu Kor) Lam	2005 - 2010	Roche	
Chong Shou	2005 - 2011	IBM	
Raymond Auerbach	2007 - 2012	US Government	
Rebecca Robilotto	2007 - 2012	Allbirds	
Xinmeng (Jasmine) Mu	2007 - 2012	Pfizer Inc.	
Jing (Jane) Leng	2009 - 2014	Illumina	
Yao Fu	2011 - 2015	Illumina	
Jieming Chen	2011 - 2016	Genentech Inc.	
Bo Wang	2013 - 2019	McKinsey	
Mengting Gu	2013 - 2019	Visa Research	

Main Scientific Publications

(See footnotes at end of the publication section)

-- 2019 --

- JS Johnson, D Spakowicz, BY Hong, LM Petersen, P Demkowicz, L Chen, SR Leopold, BM Hanson, HO Agresta, M Gerstein, E Sodergren, GM Weinstock (2019). "Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis." *Nat Commun.* 10: 5029
- J Warrell, MB Gerstein (2019). "Hierarchical PAC-Bayes Bounds via Deep Probabilistic Programming." Bayesian Deep Learning Workshop at NeurIPS.
- M Koupenova, E Mick, HA Corkrey, A Singh, SE Tanriverdi, O Vitseva, D Levy, AM Keeler, M Ezzaty Mirhashemi, MK ElMallah, M Gerstein, J Rozowsky, K Tanriverdi, JE Freedman (2019). "Pollen-derived RNAs Are Found in the Human Circulation." *iScience* 19: 916-926.
- S Lou, KA Cotter, T Li, J Liang, H Mohsen, J Liu, J Zhang, S Cohen, J Xu, H Yu, MA Rubin, M Gerstein (2019). "GRAM: A GeneRALized Model to predict the molecular effect of a non-coding variant in a cell-type specific manner." *PLoS Genet* 15: e1007860.
- S Kumar, D Clarke, MB Gerstein (2019). "Leveraging protein dynamics to identify cancer mutational hotspots using 3D structures." *Proc Natl Acad Sci U S A* 116: 18962-18970.
- D Greenbaum, M Gerstein (2019). "Sharing data." *Science* 365:764.
- FC Navarro, J Hoops, L Bellfy, E Cerveira, Q Zhu, C Zhang, C Lee, MB Gerstein (2019). "TeXP: Deconvolving the effects of pervasive and autonomous transcription of transposable elements." *PLoS Comput Biol* 15: e1007293.
- D Polioudakis, L de la Torre-Ubieta, J Langerman, AG Elkins, X Shi, JL Stein, CK Vuong, S Nichterwitz, M Gevorgian, CK Opland, D Lu, W Connell, EK Ruzzo, JK Lowe, T Hadzic, FI Hinz, S Sabri, WE Lowry, MB Gerstein, K Plath, DH Geschwind (2019). "A Single-Cell Transcriptomic Atlas of Human Neocortical Development during Mid-gestation." *Neuron* 103: 785-801e8.
- B Wang, C Yan, S Lou, P Emani, B Li, M Xu, X Kong, W Meyerson, YT Yang, D Lee, M Gerstein (2019). "Building a Hybrid Physical-Statistical Classifier for Predicting the Effect of Variants Related to Protein-Drug Interactions." *Structure* 27: 1469-1481e3.
- FCP Navarro, H Mohsen, C Yan, S Li, M Gu, W Meyerson, M Gerstein (2019). "Genomics and data science: an application within an umbrella." *Genome Biol* 20: 109.
- MJP Chaisson, AD Sanders, X Zhao, A Malhotra, D Porubsky, T Rausch, EJ Gardner, OL Rodriguez, L Guo, RL Collins, X Fan, J Wen, RE Handsaker, S Fairley, ZN Kronenberg, X Kong, F Hormozdiari, D Lee, AM Wenger, AR Hastie, D Antaki, T Anantharaman, PA Audano, H Brand, S Cantsilieris, H Cao, E Cerveira, C Chen, X Chen, CS Chin, Z Chong, NT Chuang, CC Lambert, DM Church, L Clarke, A Farrell, J Flores, T Galeev, DU Gorkin, M Gujral, V Guryev, WH Heaton, J Korlach, S Kumar, JY Kwon, ET Lam, JE Lee, J Lee, WP Lee, SP Lee, S Li, P Marks, K Viaud-Martinez, S Meiers, KM Munson, FCP Navarro, BJ Nelson, C Nodzak, A Noor, S Kyriazopoulou-Panagiotopoulou, AWC Pang, Y Qiu, G Rosanio, M Ryan, A Stutz, DCJ Spierings, A Ward, AE Welch, M Xiao, W Xu, C Zhang, Q Zhu, X Zheng-Bradley, E Lowy, S Yakneen, S McCarroll, G Jun, L Ding, CL Koh, B Ren, P Flicek, K Chen, MB Gerstein, PY Kwok, PM Lansdorp, GT Marth, J Sebat, X Shi, A Bashir, K Ye, SE Devine, ME Talkowski, RE Mills, T Marschall, JO Korbel, EE Eichler, C Lee (2019). "Multi-platform discovery of haplotype-resolved structural variation in human genomes." *Nat Commun* 10: 1784.

- J Rozowsky, RR Kitchen, JJ Park, TR Galeev, J Diao, J Warrell, W Thistlethwaite, SL Subramanian, A Milosavljevic, M Gerstein (2019). "exceRpt: A Comprehensive Analytic Platform for Extracellular RNA Profiling." *Cell Syst* 8: 352-357e3.
- OD Murillo, W Thistlethwaite, J Rozowsky, SL Subramanian, R Lucero, N Shah, AR Jackson, S Srinivasan, A Chung, CD Laurent, RR Kitchen, T Galeev, J Warrell, JA Diao, JA Welsh, K Hanspers, A Riutta, S Burgstaller-Muehlbacher, RV Shah, A Yeri, LM Jenkins, ME Ahsen, C Cordon-Cardo, N Dogra, SM Gifford, JT Smith, G Stolovitzky, AK Tewari, BH Wunsch, KK Yadav, KM Danielson, J Filant, C Moeller, P Nejad, A Paul, B Simonson, DK Wong, X Zhang, L Balaj, R Gandhi, AK Sood, RP Alexander, L Wang, C Wu, DTW Wong, DJ Galas, K Van Keuren-Jensen, T Patel, JC Jones, S Das, KH Cheung, AR Pico, AI Su, RL Raffai, LC Laurent, ME Roth, MB Gerstein, A Milosavljevic (2019). "exRNA Atlas Analysis Reveals Distinct Extracellular RNA Cargo Types and Their Carriers Present across Human Biofluids." *Cell* 177: 463-477e15.
- S Das, Extracellular RNA Communication Consortium, KM Ansel, M Bitzer, XO Breakefield, A Charest, DJ Galas, MB Gerstein, M Gupta, A Milosavljevic, MT McManus, T Patel, RL Raffai, J Rozowsky, ME Roth, JA Saugstad, K Van Keuren-Jensen, AM Weaver, LC Laurent (2019). "The Extracellular RNA Communication Consortium: Establishing Foundational Knowledge and Technologies for Extracellular RNA Research." *Cell* 177: 231-242.
- EY Ho, Q Cao, M Gu, RW Chan, Q Wu, M Gerstein, KY Yip (2019). "Shaping the nebulous enhancer in the era of high-throughput assays and genome editing." *Brief Bioinform.* bbz030
- GG Yardmc, H Ozadam, MEG Sauria, O Ursu, KK 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 (2019). "Measuring the reproducibility and quality of Hi-C data." *Genome Biol* 20: 57.
- A Moro, TP Driscoll, LC Boraas, W Armero, DM Kasper, N Baeyens, C Jouy, V Mallikarjun, J Swift, SJ Ahn, D Lee, J Zhang, M Gu, M Gerstein, M Schwartz, S Nicoli (2019). "MicroRNA-dependent regulation of biomechanical genes establishes tissue stiffness homeostasis." *Nat Cell Biol* 21: 348-358.
- JE Posey, AH O'Donnell-Luria, JX Chong, T Harel, SN Jhangiani, ZH Coban Akdemir, S Buyske, D Pehlivan, CMB Carvalho, S Baxter, N Sobreira, P Liu, N Wu, JA Rosenfeld, S Kumar, D Avramopoulos, JJ White, KF Doheny, PD Witmer, C Boehm, VR Sutton, DM Muzny, E Boerwinkle, M Gunel, DA Nickerson, S Mane, DG MacArthur, RA Gibbs, A Hamosh, RP Lifton, TC Matise, HL Rehm, M Gerstein, MJ Bamshad, D Valle, JR Lupski, Centers for Mendelian Genomics (2019). "Insights into genetics, human biology and disease gleaned from family based genomic studies." *Genet Med* 21: 798-812.
- S Li, MB Gerstein (2019). "Next-Generation Sequencing to Diagnose Suspected Genetic Disorders." *N Engl J Med* 380: 200.

-- 2018 --

- PsychENCODE Consortium (2018). "Revealing the brain's molecular architecture." *Science* 362: 1262-1263.
- D Wang, S Liu, J Warrell, H Won, X Shi, FCP Navarro, D Clarke, M Gu, P Emani, YT Yang, M Xu, MJ Gandal, S Lou, J Zhang, JJ Park, C Yan, SK Rhie, K Manakongtreecheep, H Zhou, A Nathan, M Peters, E Mattei, D Fitzgerald, T Brunetti, J Moore, Y Jiang, K Girdhar, GE Hoffman, S Kalayci, ZH Gumus, GE Crawford, PsychENCODE Consortium, P Roussos, S Akbarian, AE Jaffe, KP White, Z Weng, N Sestan, DH Geschwind, JA Knowles, MB Gerstein (2018). "Comprehensive functional genomic resource and integrative model for the human brain." *Science* 362: eaat8464.

- MJ Gandal, P Zhang, E Hadjimichael, RL Walker, C Chen, S Liu, H Won, H van Bakel, M Varghese, Y Wang, AW Shieh, J Haney, S Parhami, J Belmont, M Kim, P Moran Losada, Z Khan, J Mleczko, Y Xia, R Dai, D Wang, YT Yang, M Xu, K Fish, PR Hof, J Warrell, D Fitzgerald, K White, AE Jaffe, PsychENCODE Consortium, MA Peters, M Gerstein, C Liu, LM Iakoucheva, D Pinto, DH Geschwind (2018). "Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder." *Science* 362.
- M Li, G Santpere, Y Imamura Kawasawa, OV Evgrafov, FO Gulden, S Pochareddy, SM Sunkin, Z Li, Y Shin, Y Zhu, AMM Sousa, DM Werling, RR Kitchen, HJ Kang, M Pletikos, J Choi, S Muchnik, X Xu, D Wang, B Lorente-Galdos, S Liu, P Giusti-Rodriguez, H Won, CA de Leeuw, AF Pardinas, BrainSpan Consortium, PsychENCODE Consortium, PsychENCODE Developmental Subgroup, M Hu, F Jin, Y Li, MJ Owen, MC O'Donovan, JTR Walters, D Posthuma, P Levitt, DR Weinberger, TM Hyde, JE Kleinman, DH Geschwind, MJ Hawrylycz, MW State, SJ Sanders, PF Sullivan, MB Gerstein, ES Lein, JA Knowles, N Sestan (2018). "Integrative functional genomic analysis of human brain development and neuropsychiatric risks." *Science* 362: 1264.
- A Amiri, G Coppola, S Scuderi, F Wu, T Roychowdhury, F Liu, S Pochareddy, Y Shin, A Safi, L Song, Y Zhu, AMM Sousa, PsychENCODE Consortium, M Gerstein, GE Crawford, N Sestan, A Abyzov, FM Vaccarino (2018). "Transcriptome and epigenome landscape of human cortical development modeled in organoids." *Science* 362.
- X Kong, M Gerstein (2018). "Text mining systems biology: Turning the microscope back on the observer" *Current Opinion in Systems Biology* 11:117-122.
- W Shi, CKY Ng, RS Lim, T Jiang, S Kumar, X Li, VB Wali, S Piscuoglio, MB Gerstein, AB Chagpar, B Weigelt, L Pusztai, JS Reis-Filho, C Hatzis (2018). "Reliability of Whole-Exome Sequencing for Assessing Intratumor Genetic Heterogeneity." *Cell Rep* 25: 1446-1457.
- A Frankish, M Diekhans, AM Ferreira, R Johnson, I Jungreis, J Loveland, JM Mudge, C Sisu, J Wright, J Armstrong, I Barnes, A Berry, A Bignell, S Carbonell Sala, J Chrast, F Cunningham, T Di Domenico, S Donaldson, IT Fiddes, C Garcia Giron, JM Gonzalez, T Grego, M Hardy, T Hourlier, T Hunt, OG Izuogu, J Lagarde, FJ Martin, L Martinez, S Mohanan, P Muir, FCP Navarro, A Parker, B Pei, F Pozo, M Ruffier, BM Schmitt, E Stapleton, MM Suner, I Sycheva, B Uszczyńska-Ratajczak, J Xu, A Yates, D Zerbino, Y Zhang, B Aken, JS Choudhary, M Gerstein, R Guigo, TJP Hubbard, M Kellis, B Paten, A Raymond, ML Tress, P Flicek (2018). "GENCODE reference annotation for the human and mouse genomes." *Nucleic Acids Res.*
- J Lilue, AG Doran, IT Fiddes, M Abrudan, J Armstrong, R Bennett, W Chow, J Collins, S Collins, A Czechanski, P Danecek, M Diekhans, DD Dolle, M Dunn, R Durbin, D Earl, A Ferguson-Smith, P Flicek, J Flint, A Frankish, B Fu, M Gerstein, J Gilbert, L Goodstadt, J Harrow, K Howe, X Ibarra-Soria, M Kolmogorov, CJ Lelliott, DW Logan, J Loveland, CE Mathews, R Mott, P Muir, S Nachtweide, FCP Navarro, DT Odom, N Park, S Pelan, SK Pham, M Quail, L Reinholdt, L Romoth, L Shirley, C Sisu, M Sjoberg-Herrera, M Stanke, C Steward, M Thomas, G Threadgold, D Thybert, J Torrance, K Wong, J Wood, B Yalcin, F Yang, DJ Adams, B Paten, TM Keane (2018). "Sixteen diverse laboratory mouse reference genomes define strain-specific haplotypes and novel functional loci." *Nat Genet* 50: 1574-1583.
- D Greenbaum, M Gerstein (2018). "What's next for humanity?" *Science* 362 (6415):648.
- D Greenbaum, M Gerstein (2018). "Human History, Human Genomes" *Cell* 174:1043-1044.
- V Onuchic, E Lurie, I Carrero, P Pawliczek, RY Patel, J Rozowsky, T Galeev, Z Huang, RC Altshuler, Z Zhang, RA Harris, C Coarfa, L Ashmore, JW Bertol, WD Fakhouri, F Yu, M Kellis, M Gerstein, A Milosavljevic (2018). "Allele-specific epigenome maps reveal sequence-dependent stochastic switching at regulatory loci." *Science* 361.

- BC Carlyle, RR Kitchen, J Zhang, RS Wilson, TT Lam, JS Rozowsky, KR Williams, N Sestan, MB Gerstein, AC Nairn (2018). "Isoform-Level Interpretation of High-Throughput Proteomics Data Enabled by Deep Integration with RNA-seq." *J Proteome Res* 17: 3431-3444.
- AP Arkin, RW Cottingham, CS Henry, NL Harris, RL Stevens, S Maslov, P Dehal, D Ware, F Perez, S Canon, MW Sneddon, ML Henderson, WJ Riehl, D Murphy-Olson, SY Chan, RT Kamimura, S Kumari, MM Drake, TS Brettin, EM Glass, D Chivian, D Gunter, DJ Weston, BH Allen, J Baumohl, AA Best, B Bowen, SE Brenner, CC Bun, JM Chandonia, JM Chia, R Colasanti, N Conrad, JJ Davis, BH Davison, M DeJongh, S Devoid, E Dietrich, I Dubchak, JN Edirisinghe, G Fang, JP Faria, PM Frybarger, W Gerlach, M Gerstein, A Greiner, J Gurtowski, HL Haun, F He, R Jain, MP Joachimiak, KP Keegan, S Kondo, V Kumar, ML Land, F Meyer, M Mills, PS Novichkov, T Oh, GJ Olsen, R Olson, B Parrello, S Pasternak, E Pearson, SS Poon, GA Price, S Ramakrishnan, P Ranjan, PC Ronald, MC Schatz, SMD Seaver, M Shukla, RA Sutormin, MH Syed, J Thomason, NL Tintle, D Wang, F Xia, H Yoo, S Yoo, D Yu (2018). "KBase: The United States Department of Energy Systems Biology Knowledgebase." *Nat Biotechnol* 36: 566-569.
- A Harmanci, M Gerstein (2018). "Analysis of sensitive information leakage in functional genomics signal profiles through genomic deletions" *Nat Commun* 9: 2453.
- KW Barber, P Muir, RV Szeligowski, S Rogulina, M Gerstein, JR Sampson, FJ Isaacs, J Rinehart (2018). "Encoding human serine phosphopeptides in bacteria for proteome-wide identification of phosphorylation-dependent interactions." *Nat Biotechnol* 36: 638-644.
- P McGillivray, D Clarke, W Meyerson, J Zhang, D Lee, M Gu, S Kumar, H Zhou, MB Gerstein (2018). "Network Analysis as a Grand Unifier in Biomedical Data Science" *Annual Review of Biomedical Data Science* Vol. 1.
- D Thybert, M Roller, FCP Navarro, I Fiddes, I Streeter, C Feig, D Martin-Galvez, M Kolmogorov, V Janousek, W Akanni, B Aken, S Aldridge, V Chakrapani, W Chow, L Clarke, C Cummins, A Doran, M Dunn, L Goodstadt, K Howe, M Howell, AA Josselin, RC Karn, CM Laukaitis, L Jingtao, F Martin, M Muffato, S Nachtweide, MA Quail, C Sisu, M Stanke, K Stefflova, C Van Oosterhout, F Veyrunes, B Ward, F Yang, G Yazdanifar, A Zadissa, DJ Adams, A Brazma, M Gerstein, B Paten, S Pham, TM Keane, DT Odom, P Flicek (2018). "Repeat associated mechanisms of genome evolution and function revealed by the *Mus caroli*" *Genome Res* 28: 448-459.
- P McGillivray, R Ault, M Pawashe, R Kitchen, S Balasubramanian, M Gerstein (2018). "A comprehensive catalog of predicted functional upstream open reading frames in humans." *Nucleic Acids Res* 46: 3326-3338.
- T Becker, WP Lee, J Leone, Q Zhu, C Zhang, S Liu, J Sargent, K Shanker, A Mil-Homens, E Cerveira, M Ryan, J Cha, FCP Navarro, T Galeev, M Gerstein, RE Mills, DG Shin, C Lee, A Malhotra (2018). "FusorSV: an algorithm for optimally combining data from multiple structural variation detection methods." *Genome Biol* 19: 38.
- MJ Gandal, JR Haney, NN Parikshak, V Leppa, G Ramaswami, C Hartl, AJ Schork, V Appadurai, A Buil, TM Werge, C Liu, KP White, CommonMind Consortium, PsychENCODE Consortium, iPSYCH-BROAD Working Group, S Horvath, DH Geschwind (2018). "Shared molecular neuropathology across major psychiatric disorders parallels polygenic overlap." *Science* 359: 693-697.
- MB Gerstein, FCP Navarro (2018). "Gene names can confound most-searched listings." *Nature* 553: 405.
- BD Piening, W Zhou, K Contrepolis, H Rost, GJ Gu Urban, T Mishra, BM Hanson, EJ Bautista, S Leopold, CY Yeh, D Spakowicz, I Banerjee, C Chen, K Kukurba, D Perelman, C Craig, E Colbert, D Salins, S Rego, S Lee, C Zhang, J Wheeler, MR Sailani, L Liang, C Abbott, M Gerstein, A Mardinoglu, U Smith, DL Rubin, S Pitteri, E Sodergren, TL McLaughlin, GM Weinstock, MP Snyder (2018). "Integrative Personal Omics Profiles during Periods of Weight Gain and Loss." *Cell Syst* 6: 157-170e8.

-- 2017 --

- KE Kaczor-Urbanowicz, Y Kim, F Li, T Galeev, RR Kitchen, M Gerstein, K Koyano, SH Jeong, X Wang, D Elashoff, SY Kang, SM Kim, K Kim, S Kim, D Chia, X Xiao, J Rozowsky, DTW Wong (2017). "Novel approaches for bioinformatic analysis of salivary RNA sequencing data for development." *Bioinformatics* 34: 1-8.
- MM Kudron, A Victorsen, L Gevirtzman, LW Hillier, WW Fisher, D Vafeados, M Kirkey, AS Hammonds, J Gersch, H Ammouri, ML Wall, J Moran, D Steffen, M Szykarek, S Seabrook-Sturgis, N Jameel, M Kadaba, J Patton, R Terrell, M Corson, TJ Durham, S Park, S Samanta, M Han, J Xu, KK Yan, SE Celniker, KP White, L Ma, M Gerstein, V Reinke, RH Waterston (2017). "The ModERN Resource: Genome-Wide Binding Profiles for Hundreds of Drosophila and Caenorhabditis elegans Transcription Factors." *Genetics* 208: 937-949.
- BC Carlyle, RR Kitchen, JE Kanyo, EZ Voss, M Pletikos, AMM Sousa, TT Lam, MB Gerstein, N Sestan, AC Nairn (2017). "A multiregional proteomic survey of the postnatal human brain." *Nat Neurosci* 20: 1787-1795.
- AMM Sousa, Y Zhu, MA Raghanti, RR Kitchen, M Onorati, ATN Tebbenkamp, B Stutz, KA Meyer, M Li, YI Kawasawa, F Liu, RG Perez, M Mele, T Carvalho, M Skarica, FO Gulden, M Pletikos, A Shibata, AR Stephenson, MK Edler, JJ Ely, JD Elsworth, TL Horvath, PR Hof, TM Hyde, JE Kleinman, DR Weinberger, M Reimers, RP Lifton, SM Mane, JP Noonan, MW State, ES Lein, JA Knowles, T Marques-Bonet, CC Sherwood, MB Gerstein, N Sestan (2017). "Molecular and cellular reorganization of neural circuits in the human lineage." *Science* 358: 1027-1032.
- L Lochovsky, J Zhang, M Gerstein (2017). "MOAT: efficient detection of highly mutated regions with the Mutations Overburdening Annotations Tool." *Bioinformatics* 34: 1031-1033.
- Q Cao, C Anyansi, X Hu, L Xu, L Xiong, W Tang, MTS Mok, C Cheng, X Fan, M Gerstein, ASL Cheng, KY Yip (2017). "Reconstruction of enhancer-target networks in 935 samples of human primary cells, tissues and cell lines." *Nat Genet* 49: 1428-1436.
- S Balasubramanian, Y Fu, M Pawashe, P McGillivray, M Jin, J Liu, KJ Karczewski, DG MacArthur, M Gerstein (2017). "Using ALoFT to determine the impact of putative loss-of-function variants in protein-coding genes." *Nat Commun* 8: 382.
- KK Yan, S Lou, M Gerstein (2017). "MrTADFinder: A network modularity based approach to identify topologically associating domains in multiple resolutions." *PLoS Comput Biol* 13: e1005647.
- SME Sahraeian, M Mohiyuddin, R Sebra, H Tilgner, PT Afshar, KF Au, N Bani Asadi, MB Gerstein, WH Wong, MP Snyder, E Schadt, HYK Lam (2017). "Gaining comprehensive biological insight into the transcriptome by performing a broad-spectrum RNA-seq analysis." *Nat Commun* 8: 59.
- Y Zhang, S Li, A Abyzov, MB Gerstein (2017). "Landscape and variation of novel retroduplications in 26 human populations." *PLoS Comput Biol* 13: e1005567.
- S Kumar, M Gerstein (2017). "Cancer genomics: Less is more in the hunt for driver mutations." *Nature* 547: 40-41.
- P Dhingra, Y Fu, M Gerstein, E Khurana (2017). "Using FunSeq2 for Coding and Non-Coding Variant Annotation and Prioritization." *Curr Protoc Bioinformatics* 57: 15111-15117.

P Alves, S Liu, D Wang, M Gerstein (2017). "Multiple-Swarm Ensembles: Improving the Predictive Power and Robustness of Predictive Models and Its Use in Computational Biology." *IEEE/ACM Trans Comput Biol Bioinform* 15: 926-933.

V Despic, M Dejung, M Gu, J Krishnan, J Zhang, L Herzel, K Straube, MB Gerstein, F Butter, KM Neugebauer (2017). "Dynamic RNA-protein interactions underlie the zebrafish maternal-to-zygotic transition." *Genome Res* 27: 1184-1194.

D Greenbaum, J Rozowsky, V Stodden, M Gerstein (2017). "Structuring supplemental materials in support of reproducibility." *Genome Biol* 18: 64.

KK Yan, GG Yardimci, C Yan, WS Noble, M Gerstein (2017). "HiC-spector: a matrix library for spectral and reproducibility analysis of Hi-C contact maps." *Bioinformatics* 33: 2199-2201.

S Li, BM Shuch, MB Gerstein (2017). "Whole-genome analysis of papillary kidney cancer finds significant noncoding alterations." *PLoS Genet* 13: e1006685.

DM Kasper, A Moro, E Ristori, A Narayanan, G Hill-Teran, E Fleming, M Moreno-Mateos, CE Vejnar, J Zhang, D Lee, M Gu, M Gerstein, A Giraldez, S Nicoli (2017). "MicroRNAs Establish Uniform Traits during the Architecture of Vertebrate Embryos." *Dev Cell* 40: 552-565e5.

E Mick, R Shah, K Tanriverdi, V Murthy, M Gerstein, J Rozowsky, R Kitchen, MG Larson, D Levy, JE Freedman (2017). "Stroke and Circulating Extracellular RNAs." *Stroke* 48: 828-834.

A Abyzov, L Tomasini, B Zhou, N Vasmatzis, G Coppola, M Amenduni, R Pattni, M Wilson, M Gerstein, S Weissman, AE Urban, FM Vaccarino (2017). "One thousand somatic SNVs per skin fibroblast cell set baseline of mosaic mutational load with patterns that suggest proliferative origin." *Genome Res* 27: 512-523.

-- 2016 --

J Chen, B Wang, L Regan, M Gerstein (2016). "Intensification: A Resource for Amplifying Population-Genetic Signals with Protein Repeats." *J Mol Biol* 429: 435-445.

S Kumar, D Clarke, M Gerstein (2016). "Localized structural frustration for evaluating the impact of sequence variants." *Nucleic Acids Res* 44: 10062-10073.

D Wang, F He, S Maslov, M Gerstein (2016). "DREISS: Using State-Space Models to Infer the Dynamics of Gene Expression Driven by External and Internal Regulatory Networks." *PLoS Comput Biol* 12: e1005146.

D Greenbaum, M Gerstein (2016). "Opinion: GMOs Are Not "Frankenfoods"" *The Scientist* (30 Aug

CC Yang, EH Andrews, MH Chen, WY Wang, JJ Chen, M Gerstein, CC Liu, C Cheng (2016). "iTAR: a web server for identifying target genes of transcription factors using ChIP-seq or ChIP-chip data." *BMC Genomics* 17: 632.

SW Choo, M Rayko, TK Tan, R Hari, A Komissarov, WY Wee, AA Yurchenko, S Kliver, G Tamazian, A Antunes, RK Wilson, WC Warren, KP Koepfli, P Minx, K Krashennikova, A Kotze, DL Dalton, E Vermaak, IC Paterson, P Dobrynin, FT Sitam, JJ Rovie-Ryan, WE Johnson, AM Yusoff, SJ Luo, KV Karuppanan, G Fang, D Zheng, MB Gerstein, L Lipovich, SJ O'Brien, GJ Wong (2016). "Pangolin genomes and the evolution of mammalian scales and immunity." *Genome Res* 26: 1312-1322.

- R Shah, K Tanriverdi, D Levy, M Larson, M Gerstein, E Mick, J Rozowsky, R Kitchen, V Murthy, E Mikalev, JE Freedman (2016). "Discordant Expression of Circulating microRNA from Cellular and Extracellular Sources." *PLoS One* 11: e0153691.
- JE Freedman, M Gerstein, E Mick, J Rozowsky, D Levy, R Kitchen, S Das, R Shah, K Danielson, L Beaulieu, FC Navarro, Y Wang, TR Galeev, A Holman, RY Kwong, V Murthy, SE Tanriverdi, M Koupenova-Zamor, E Mikhalev, K Tanriverdi (2016). "Diverse human extracellular RNAs are widely detected in human plasma." *Nat Commun* 7: 11106.
- KH Cheung, S Keerthikumar, P Roncaglia, SL Subramanian, ME Roth, M Samuel, S Anand, L Gangoda, S Gould, R Alexander, D Galas, MB Gerstein, AF Hill, RR Kitchen, J Lotvall, T Patel, DC Procaccini, P Quesenberry, J Rozowsky, RL Raffai, A Shypitsyna, AI Su, C Thery, K Vickers, MH Wauben, S Mathivanan, A Milosavljevic, LC Laurent (2016). "Extending gene ontology in the context of extracellular RNA and vesicle communication." *J Biomed Semantics* 7: 19.
- J Chen, J Rozowsky, TR Galeev, A Harmanci, R Kitchen, J Bedford, A Abyzov, Y Kong, L Regan, M Gerstein (2016). "A uniform survey of allele-specific binding and expression over 1000-Genomes-Project individuals." *Nat Commun* 7: 11101.
- D Clarke, A Sethi, S Li, S Kumar, RWF Chang, J Chen, M Gerstein (2016). "Identifying Allosteric Hotspots with Dynamics: Application to Inter- and Intra-species Conservation." *Structure* 24: 826-837.
- KK Yan, D Wang, A Sethi, P Muir, R Kitchen, C Cheng, M Gerstein (2016). "Cross-Disciplinary Network Comparison: Matchmaking Between Hairballs." *Cell Syst* 2: 147-157.
- D Greenbaum, M Gerstein (2016). "Who Owns Your DNA?" *Cell* 165:257-258.
- F He, S Yoo, D Wang, S Kumari, M Gerstein, D Ware, S Maslov (2016). "Large-scale atlas of microarray data reveals the distinct expression landscape of different tissues in Arabidopsis." *Plant J* 86: 472-80.
- P Muir, S Li, S Lou, D Wang, DJ Spakowicz, L Salichos, J Zhang, GM Weinstock, F Isaacs, J Rozowsky, M Gerstein (2016). "The real cost of sequencing: scaling computation to keep pace with data generation." *Genome Biol* 17: 53.
- D Wang, KK Yan, J Rozowsky, E Pan, M Gerstein (2016). "Temporal Dynamics of Collaborative Networks in Large Scientific Consortia." *Trends Genet* 32: 251-253.
- A Harmanci, M Gerstein (2016). "Quantification of private information leakage from phenotype-genotype data: linking attacks." *Nat Methods* 13: 251-6.
- E Khurana, Y Fu, D Chakravarty, F Demichelis, MA Rubin, M Gerstein (2016). "Role of non-coding sequence variants in cancer." *Nat Rev Genet* 17: 93-108.
- A Abyzov, S Li, MB Gerstein (2016). "Understanding genome structural variations." *Oncotarget* 7: 7370-1.

-- 2015 --

Cancer Genome Atlas Research Network. "The Molecular Taxonomy of Primary Prostate Cancer." *Cell* 163: 1011-25.

S Purushotham, MR Min, CJ Kuo, M Gerstein (2015). " Knowledge Based Factorized High Order Sparse Learning Models." NIPS Workshop on Machine Learning in Computational Biology.

- A Abyzov, S Li, MB Gerstein (2015). "Understanding genome structural variations." *Oncotarget* .
- A Sethi, D Clarke, J Chen, S Kumar, TR Galeev, L Regan, M Gerstein (2015). "Reads meet rotamers: structural biology in the age of deep sequencing." *Curr Opin Struct Biol* 35: 125-34.
- S Akbarian, C Liu, JA Knowles, FM Vaccarino, PJ Farnham, GE Crawford, AE Jaffe, D Pinto, S Dracheva, DH Geschwind, J Mill, AC Nairn, A Abyzov, S Pochareddy, S Prabhakar, S Weissman, PF Sullivan, MW State, Z Weng, MA Peters, KP White, MB Gerstein, A Amiri, C Armoskus, AE Ashley-Koch, T Bae, A Beckel-Mitchener, BP Berman, GA Coetzee, G Coppola, N Francoeur, M Fromer, R Gao, K Grennan, J Herstein, DH Kavanagh, NA Ivanov, Y Jiang, RR Kitchen, A Kozlenkov, M Kundakovic, M Li, Z Li, S Liu, LM Mangravite, E Mattei, E Markenscoff-Papadimitriou, FC Navarro, N North, L Omberg, D Panchision, N Parikshak, J Poschmann, AJ Price, M Purcaro, TE Reddy, P Roussos, S Schreiner, S Scuderi, R Sebra, M Shibata, AW Shieh, M Skarica, W Sun, V Swarup, A Thomas, J Tsuji, H van Bakel, D Wang, Y Wang, K Wang, DM Werling, AJ Willsey, H Witt, H Won, CC Wong, GA Wray, EY Wu, X Xu, L Yao, G Senthil, T Lehner, P Sklar, N Sestan (2015). "The PsychENCODE project." *Nat Neurosci* 18: 1707-12.
- D Greenbaum, M Gerstein (2015). "Illuminating the Genome's Dark Matter," *Cell* 163:1047-1048.
- Cancer Genome Atlas Research Network, WM Linehan, PT Spellman, CJ Ricketts, CJ Creighton, SS Fei, C Davis, DA Wheeler, BA Murray, L Schmidt, CD Vocke, M Peto, AA Al Mamun, E Shinbrot, A Sethi, S Brooks, WK Rathmell, AN Brooks, KA Hoadley, AG Robertson, D Brooks, R Bowlby, S Sadeghi, H Shen, DJ Weisenberger, M Bootwalla, SB Baylin, PW Laird, AD Cherniack, G Saksena, S Haake, J Li, H Liang, Y Lu, GB Mills, R Akbani, MD Leiserson, BJ Raphael, P Anur, D Bottaro, L Albiges, N Barnabas, TK Choueiri, B Czerniak, AK Godwin, AA Hakimi, TH Ho, J Hsieh, M Ittmann, WY Kim, B Krishnan, MJ Merino, KR Mills Shaw, VE Reuter, E Reznik, CS Shelley, B Shuch, S Signoretti, R Srinivasan, P Tamboli, G Thomas, S Tickoo, K Burnett, D Crain, J Gardner, K Lau, D Mallery, S Morris, JD Paulauskis, RJ Penny, C Shelton, WT Shelton, M Sherman, E Thompson, P Yena, MT Avedon, J Bowen, JM Gastier-Foster, M Gerken, KM Leraas, TM Lichtenberg, NC Ramirez, T Santos, L Wise, E Zmuda, JA Demchok, I Felau, CM Hutter, M Sheth, HJ Sofia, R Tarnuzzer, Z Wang, L Yang, JC Zenklusen, J Zhang, B Ayala, J Baboud, S Chudamani, J Liu, L Lolla, R Naresh, T Pihl, Q Sun, Y Wan, Y Wu, A Ally, M Balasundaram, S Balu, R Beroukhi, T Bodenheimer, C Buhay, YS Butterfield, R Carlsen, SL Carter, H Chao, E Chuah, A Clarke, KR Covington, M Dahdouli, N Dewal, N Dhalla, HV Doddapaneni, JA Drummond, SB Gabriel, RA Gibbs, R Guin, W Hale, A Hawes, DN Hayes, RA Holt, AP Hoyle, SR Jefferys, SJ Jones, CD Jones, D Kalra, C Kovar, L Lewis, J Li, Y Ma, MA Marra, M Mayo, S Meng, M Meyerson, PA Mieczkowski, RA Moore, D Morton, LE Mose, AJ Mungall, D Muzny, JS Parker, CM Perou, J Roach, JE Schein, SE Schumacher, Y Shi, JV Simons, P Sipahimalani, T Skelly, MG Soloway, C Sougnez, A Tam, D Tan, N Thiessen, U Veluvolu, M Wang, MD Wilkerson, T Wong, J Wu, L Xi, J Zhou, J Bedford, F Chen, Y Fu, M Gerstein, D Haussler, K Kasaian, P Lai, S Ling, A Radenbaugh, D Van Den Berg, JN Weinstein, J Zhu, M Albert, I Alexopoulou, JJ Andersen, JT Auman, J Bartlett, S Bastacky, J Bergsten, ML Blute, L Boice, RJ Bollag, J Boyd, E Castle, YB Chen, JC Cheville, E Curley, B Davies, A DeVolk, R Dhir, L Dike, J Eckman, J Engel, J Harr, R Hrebinko, M Huang, L Huelsenbeck-Dill, M Iacocca, B Jacobs, M Lobis, JK Maranchie, S McMeekin, J Myers, J Nelson, J Parfitt, A Parwani, N Petrelli, B Rabeno, S Roy, AL Salner, J Slaton, M Stanton, RH Thompson, L Thorne, K Tucker, PM Weinberger, C Winemiller, LA Zach, R Zuna (2015). "Comprehensive Molecular Characterization of Papillary Renal-Cell Carcinoma." *N Engl J Med* 374: 135-45.
- The 1000 Genomes Project Consortium (2015). "A global reference for human genetic variation." *Nature* 526: 68-74.

- PH Sudmant, T Rausch, E Gardner, R Handsaker, A Abyzov, J Huddleston, Y Zhang, K Ye, G Jun, M Fritz, M Konkel, A Malhotra, A Stütz, X Shi, F Paolo Casale, J Chen, F Hormozdiari, G Dayama, K Chen, M Malig, M Chaisson, K Walter, S Meiers, S Kashin, E Garrison, A Auton, H Lam, XJ Mu, C Alkan, D Antaki, T Bae, E Cerveira, P Chines, Z Chong, L Clarke, E Dal, L Ding, S Emery, X Fan, M Gujral, F Kahveci, J Kidd, Y Kong, E Lameijer, S McCarthy, P Flicek, R Gibbs, G Marth, C Mason, A Menelaou, D Muzny, B Nelson, A Noor, N Parrish, M Pendleton, A Quitadamo, B Raeder, E Schadt, M Romanovitch, A Schlattl, R Sebra, A Shabalina, A Untergasser, J Walker, M Wang, F Yu, C Zhang, J Zhang, X Zheng-Bradley, W Zhou, T Zichner, J Sebat, M Batzer, S McCarroll, The 1000 Genomes Project Consortium, R Mills, M Gerstein, A Bashir, O Stegle, S Devine, C Lee, E Eichler, JO Korbel (2015). "An integrated map of structural variation in 2,504 human genomes." *Nature* 526: 75–81.
- JC Mu, P Tootoonchi Afshar, M Mohiyuddin, X Chen, J Li, N Bani Asadi, MB Gerstein, WH Wong, HY Lam (2015). "Leveraging long read sequencing from a single individual to provide a comprehensive resource for benchmarking variant calling methods." *Sci Rep* 5: 14493.
- LT Fang, PT Afshar, A Chhibber, M Mohiyuddin, Y Fan, JC Mu, G Gibeling, S Barr, NB Asadi, MB Gerstein, DC Koboldt, W Wang, WH Wong, HY Lam (2015). "An ensemble approach to accurately detect somatic mutations using SomaticSeq." *Genome Biol* 16: 197.
- EE Duffy, M Rutenberg-Schoenberg, CD Stark, RR Kitchen, MB Gerstein, MD Simon (2015). "Tracking Distinct RNA Populations Using Efficient and Reversible Covalent Chemistry." *Mol Cell* 59: 858-66.
- L Lochovsky, J Zhang, Y Fu, E Khurana, M Gerstein (2015). "LARVA: an integrative framework for large-scale analysis of recurrent variants in noncoding annotations." *Nucleic Acids Res* .
- PP Kuksa, MR Min, R Dugar, M Gerstein (2015). "High-order neural networks and kernel methods for peptide-MHC binding prediction." *Bioinformatics* .
- J Mariani, G Coppola, P Zhang, A Abyzov, L Provini, L Tomasini, M Amenduni, A Szekely, D Palejev, M Wilson, M Gerstein, EL Grigorenko, K Chawarska, KA Pelphrey, JR Howe, FM Vaccarino. (2015) "FOXG1-Dependent Dysregulation of GABA/Glutamate Neuron Differentiation in Autism Spectrum Disorders." *Cell* Jul 16;162(2):375-90.
- A Abyzov, S Li, DR Kim, M Mohiyuddin, AM Stütz, NF Parrish, XJ Mu, W Clark, K Chen, M Hurles, JO Korbel, HY Lam, C Lee, MB Gerstein (2015). "Analysis of deletion breakpoints from 1,092 humans reveals details of mutation mechanisms." *Nat Commun* 6: 7256.
- D Greenbaum, M Gerstein (2015). "The computer connection." *Science* 347: 956.
- M Mohiyuddin, JC Mu, J Li, N Bani Asadi, MB Gerstein, A Abyzov, WH Wong, HY Lam (2015). "MetaSV: an accurate and integrative structural-variant caller for next generation sequencing." *Bioinformatics* 31: 2741-4.
- D Wang, KK Yan, C Sisu, C Cheng, J Rozowsky, W Meyerson, MB Gerstein (2015). "Loregic: a method to characterize the cooperative logic of regulatory factors." *PLoS Comput Biol* 11: e1004132.
- C Cheng, E Andrews, KK Yan, M Ung, D Wang, M Gerstein (2015). "An approach for determining and measuring network hierarchy applied to comparing the phosphorlome and the regulome." *Genome Biol* 16: 63.

-- 2014 --

- JC Mu, M Mohiyuddin, J Li, N Bani Asadi, MB Gerstein, A Abyzov, WH Wong, HY Lam (2014). "VarSim: a high-fidelity simulation and validation framework for high-throughput genome sequencing with cancer applications." *Bioinformatics* 31: 1469-71.

- RR Kitchen, JS Rozowsky, MB Gerstein, AC Nairn (2014). "Decoding neuroproteomics: integrating the genome, translome and functional anatomy." *Nat Neurosci* 17: 1491-9.
- MR Min, X Ning, Y Qi, C Cheng, A Bonner, M Gerstein (2014). "Ensemble Learning Based Sparse High-Order Boltzmann Machine for Unsupervised Feature Interaction Identification." NIPS Workshop on Machine Learning in Computational Biology.
- MR Min, X Ning, C Cheng, M Gerstein (2014). "Interpretable Sparse High-Order Boltzmann Machines." *JMLR W&CP* 33: 614-622 (AISTATS 2014).
- A Harmanci, J Rozowsky, M Gerstein (2014). "MUSIC: identification of enriched regions in ChIP-Seq experiments using a mappability-corrected multiscale signal processing framework." *Genome Biol* 15: 474.
- Y Fu, Z Liu, S Lou, J Bedford, XJ Mu, KY Yip, E Khurana, M Gerstein (2014). "FunSeq2: a framework for prioritizing noncoding regulatory variants in cancer." *Genome Biol* 15: 480.
- AP Boyle, CL Araya, C Brdlik, P Cayting, C Cheng, Y Cheng, K Gardner, LW Hillier, J Janette, L Jiang, D Kasper, T Kawli, P Kheradpour, A Kundaje, JJ Li, L Ma, W Niu, EJ Rehm, J Rozowsky, M Slattery, R Spokony, R Terrell, D Vafeados, D Wang, P Weisdepp, YC Wu, D Xie, KK Yan, EA Feingold, PJ Good, MJ Pazin, H Huang, PJ Bickel, SE Brenner, V Reinke, RH Waterston, M Gerstein, KP White, M Kellis, M Snyder (2014). "Comparative analysis of regulatory information and circuits across distant species." *Nature* 512: 453-6.
- MB Gerstein, J Rozowsky, KK Yan, D Wang, C Cheng, JB Brown, CA Davis, L Hillier, C Sisu, JJ Li, B Pei, AO Harmanci, MO Duff, S Djebali, RP Alexander, BH Alver, R Auerbach, K Bell, PJ Bickel, ME Boeck, NP Boley, BW Booth, L Cherbas, P Cherbas, C Di, A Dobin, J Drenkow, B Ewing, G Fang, M Fastuca, EA Feingold, A Frankish, G Gao, PJ Good, R Guigó, A Hammonds, J Harrow, RA Hoskins, C Howald, L Hu, H Huang, TJ Hubbard, C Huynh, S Jha, D Kasper, M Kato, TC Kaufman, RR Kitchen, E Ladewig, J Lagarde, E Lai, J Leng, Z Lu, M MacCoss, G May, R McWhirter, G Merrihew, DM Miller, A Mortazavi, R Murad, B Oliver, S Olson, PJ Park, MJ Pazin, N Perrimon, D Pervouchine, V Reinke, A Reymond, G Robinson, A Samsonova, GI Saunders, F Schlesinger, A Sethi, FJ Slack, WC Spencer, MH Stoiber, P Strasbourger, A Tanzer, OA Thompson, KH Wan, G Wang, H Wang, KL Watkins, J Wen, K Wen, C Xue, L Yang, K Yip, C Zaleski, Y Zhang, H Zheng, SE Brenner, BR Graveley, SE Celniker, TR Gingeras, R Waterston (2014). "Comparative analysis of the transcriptome across distant species." *Nature* 512: 445-8.
- C Sisu, B Pei, J Leng, A Frankish, Y Zhang, S Balasubramanian, R Harte, D Wang, M Rutenberg-Schoenberg, W Clark, M Diekhans, J Rozowsky, T Hubbard, J Harrow, MB Gerstein (2014). "Comparative analysis of pseudogenes across three phyla." *Proc Natl Acad Sci U S A* 111: 13361-6.
- KK Yan, D Wang, J Rozowsky, H Zheng, C Cheng, M Gerstein (2014). "OrthoClust: an orthology-based network framework for clustering data across multiple species." *Genome Biol* 15: R100.
- DG MacArthur, TA Manolio, DP Dimmock, HL Rehm, J Shendure, GR Abecasis, DR Adams, RB Altman, SE Antonarakis, EA Ashley, JC Barrett, LG Biesecker, DF Conrad, GM Cooper, NJ Cox, MJ Daly, MB Gerstein, DB Goldstein, JN Hirschhorn, SM Leal, LA Pennacchio, JA Stamatoyannopoulos, SR Sunyaev, D Valle, BF Voight, W Winckler, C Gunter (2014). "Guidelines for investigating causality of sequence variants in human disease." *Nature* 508: 469-76.
- M Kellis, B Wold, MP Snyder, BE Bernstein, A Kundaje, GK Marinov, LD Ward, E Birney, GE Crawford, J Dekker, I Dunham, LL Elnitski, PJ Farnham, EA Feingold, M Gerstein, MC Giddings, DM Gilbert, TR Gingeras, ED Green, R Guigo, T Hubbard, J Kent, JD Lieb, RM Myers, MJ Pazin, B Ren, JA Stamatoyannopoulos, Z Weng, KP White, RC Hardison (2014). "Defining functional DNA elements in the human genome." *Proc Natl Acad Sci U S A* 111: 6131-8.

JA Miller, SL Ding, SM Sunkin, KA Smith, L Ng, A Szafer, A Ebbert, ZL Riley, JJ Royall, K Aiona, JM Arnold, C Bennet, D Bertagnolli, K Brouner, S Butler, S Caldejon, A Carey, C Cuhacian, RA Dalley, N Dee, TA Dolbeare, BA Facer, D Feng, TP Fliss, G Gee, J Goldy, L Gourley, BW Gregor, G Gu, RE Howard, JM Jochim, CL Kuan, C Lau, CK Lee, F Lee, TA Lemon, P Lesnar, B McMurray, N Mastan, N Mosqueda, T Naluai-Cecchini, NK Ngo, J Nyhus, A Oldre, E Olson, J Parente, PD Parker, SE Parry, A Stevens, M Pletikos, M Reding, K Roll, D Sandman, M Sarreal, S Shapouri, NV Shapovalova, EH Shen, N Sjoquist, CR Slaughterbeck, M Smith, AJ Sodt, D Williams, L Zöllei, B Fischl, MB Gerstein, DH Geschwind, IA Glass, MJ Hawrylycz, RF Hevner, H Huang, AR Jones, JA Knowles, P Levitt, JW Phillips, N Sestan, P Wahnoutka, C Dang, A Bernard, JG Hohmann, ES Lein (2014). "Transcriptional landscape of the prenatal human brain." *Nature* 508: 199-206.

-- 2013 --

KY Yip, C Cheng, M Gerstein (2013). "Machine learning and genome annotation: a match meant to be?" *Genome Biol* 14: 205.

E Khurana, Y Fu, J Chen, M Gerstein (2013). "Interpretation of genomic variants using a unified biological network approach." *PLoS Comput Biol* 9: e1002886.

MR Min, X Ning, C Cheng, M Gerstein (2013). "Interpretable Sparse High-Order Boltzmann Machines for Transcription Factor Interaction Identification." NIPS Workshop on Machine Learning in Computational Biology.

T Steijger, JF Abril, PG Engström, F Kokocinski, The RGASP Consortium, JF Abril, M Akerman, T Alioto, G Ambrosini, SE Antonarakis, J Behr, P Bertone, R Bohnert, P Bucher, N Cloonan, T Derrien, S Djebali, J Du, S Dudoit, PG Engström, M Gerstein, TR Gingeras, D Gonzalez, SM Grimmond, R Guigó, L Habegger, J Harrow, TJ Hubbard, C Iseli, G Jean, A Kahles, F Kokocinski, J Lagarde, J Leng, G Lefebvre, S Lewis, A Mortazavi, P Niermann, G Ratsch, A Reymond, P Ribeca, H Richard, J Rougemont, J Rozowsky, M Sammeth, A Sboner, MH Schulz, SM Searle, ND Solorzano, V Solovyev, M Stanke, T Steijger, BJ Stevenson, H Stockinger, A Valsesia, D Weese, S White, BJ Wold, J Wu, TD Wu, G Zeller, D Zerbino, MQ Zhang, TJ Hubbard, R Guigó, J Harrow, P Bertone (2013). "Assessment of transcript reconstruction methods for RNA-seq." *Nat Methods* 10:1177.

D Greenbaum, J Chen, M Gerstein (2013). "Deep Inside Champions, Just Genes?" *Science* 342: 560.

E Khurana, Y Fu, V Colonna, XJ Mu, HM Kang, T Lappalainen, A Sboner, L Lochovsky, J Chen, A Harmanci, J Das, A Abyzov, S Balasubramanian, K Beal, D Chakravarty, D Challis, Y Chen, D Clarke, L Clarke, F Cunningham, US Evani, P Flicek, R Fragoza, E Garrison, R Gibbs, ZH Gumus, J Herrero, N Kitabayashi, Y Kong, K Lage, V Liliushvili, SM Lipkin, DG MacArthur, G Marth, D Muzny, TH Pers, GR Ritchie, JA Rosenfeld, C Sisu, X Wei, M Wilson, Y Xue, F Yu, 1000 Genomes Project Consortium, ET Dermitzakis, H Yu, MA Rubin, C Tyler-Smith, M Gerstein (2013). "Integrative annotation of variants from 1092 humans: application to cancer genomics." *Science* 342: 1235587

A Abyzov, R Iskow, O Gokcumen, DW Radke, S Balasubramanian, B Pei, L Habegger, The 1000 Genomes Project Consortium, C Lee, M Gerstein (2013). "Analysis of variable retroduplications in human populations suggests coupling of retrotransposition to cell division." *Genome Res.* 12: 2042

C Cheng, Y Fu, L Shen, M Gerstein (2013). "Identification of yeast cell cycle regulated genes based on genomic features." *BMC Syst Biol* 7: 70.

Y Zhang, R Haraksingh, F Grubert, A Abyzov, M Gerstein, S Weissman, AE Urban (2013). "Child development and structural variation in the human genome." *Child Dev* 84: 34-48.

R Kittler, J Zhou, S Hua, L Ma, Y Liu, E Pendleton, C Cheng, M Gerstein, KP White (2013). "A comprehensive nuclear receptor network for breast cancer cells." *Cell Rep* 3: 538-51.

H Tilgner, D Raha, L Habegger, M Mohiuddin, M Gerstein, M Snyder (2013). "Accurate identification and analysis of human mRNA isoforms using deep long read sequencing." *G3* (Bethesda) 3: 387-97.

SB Montgomery, DL Goode, E Kvikstad, CA Albers, ZD Zhang, XJ Mu, G Ananda, B Howie, KJ Karczewski, KS Smith, V Anaya, R Richardson, J Davis, 1000 Genomes Project Consortium, DG MacArthur, A Sidow, L Duret, M Gerstein, KD Makova, J Marchini, G McVean, G Lunter (2013). "The origin, evolution, and functional impact of short insertion-deletion variants identified in 179 human genomes." *Genome Res* 23: 749-61.

WF Doolittle, P Fraser, MB Gerstein, BR Graveley, S Henikoff, C Huttenhower, A Oshlack, CP Ponting, JL Rinn, MC Schatz, J Ule, D Weigel, GM Weinstock (2013). "Sixty years of genome biology." *Genome Biol* 14: 113.

E Khurana, Y Fu, J Chen, M Gerstein (2013). "Interpretation of genomic variants using a unified biological network approach." *PLoS Comput Biol* 9: e1002886.

-- 2012 --

PC Lin, YL Chiu, S Banerjee, K Park, JM Mosquera, E Giannopoulou, P Alves, AK Tewari, MB Gerstein, H Beltran, AM Melnick, O Elemento, F Demichelis, MA Rubin (2012). "Epigenetic repression of miR-31 disrupts androgen receptor homeostasis and contributes to prostate cancer progression." *Cancer Res* 73: 1232-44.

A Abyzov, J Mariani, D Palejev, Y Zhang, MS Haney, L Tomasini, AF Ferrandino, LA Rosenberg Belmaker, A Szekely, M Wilson, A Kocabas, NE Calixto, EL Grigorenko, A Huttner, K Chawarska, S Weissman, AE Urban*, M Gerstein*, FM Vaccarino* (2012). "Somatic copy number mosaicism in human skin revealed by induced pluripotent stem cells." *Nature* 492: 438-42

1000 Genomes Project Consortium (2012). "An integrated map of genetic variation from 1,092 human genomes." *Nature* 491: 56-65.

M Gerstein (2012). "Genomics: ENCODE leads the way on big data." *Nature* 489: 208.

ENCODE Project Consortium (2012). "An integrated encyclopedia of DNA elements in the human genome." *Nature* 489: 57-74.

MB Gerstein, A Kundaje, M Hariharan, SG Landt, KK Yan, C Cheng, XJ Mu, E Khurana, J Rozowsky, R Alexander, R Min, P Alves, A Abyzov, N Addleman, N Bhardwaj, AP Boyle, P Cayting, A Charos, DZ Chen, Y Cheng, D Clarke, C Eastman, G Euskirchen, S Fietze, Y Fu, J Gertz, F Grubert, A Harmanci, P Jain, M Kasowski, P Lacroute, J Leng, J Lian, H Monahan, H O'Geen, Z Ouyang, EC Partridge, D Patacsil, F Pauli, D Raha, L Ramirez, TE Reddy, B Reed, M Shi, T Slifer, J Wang, L Wu, X Yang, KY Yip, G Zilberman-Schapira, S Batzoglou, A Sidow, PJ Farnham, RM Myers, SM Weissman, M Snyder (2012). "Architecture of the human regulatory network derived from ENCODE data." *Nature* 489: 91-100.

S Djebali, CA Davis, A Merkel, A Dobin, T Lassmann, A Mortazavi, A Tanzer, J Lagarde, W Lin, F Schlesinger, C Xue, GK Marinov, J Khatun, BA Williams, C Zaleski, J Rozowsky, M Röder, F Kokocinski, RF Abdelhamid, T Alioto, I Antoshechkin, MT Baer, NS Bar, P Batut, K Bell, I Bell, S Chakraborty, X Chen, J Chrast, J Curado, T Derrien, J Drenkow, E Dumais, J Dumais, R Duttagupta, E Falconnet, M Fastuca, K Fejes-Toth, P Ferreira, S Foissac, MJ Fullwood, H Gao, D Gonzalez, A Gordon, H Gunawardena, C Howald, S Jha, R Johnson, P Kapranov, B King, C Kingswood, OJ Luo, E Park, K Persaud, JB Preall, P Ribeca, B Risk, D Robyr, M Sammeth, L Schaffer, LH See, A Shahab, J Skancke, AM Suzuki, H Takahashi, H Tilgner, D Trout, N Walters, H Wang, J Wrobel, Y Yu, X Ruan, Y Hayashizaki, J Harrow, M Gerstein, T Hubbard, A Reymond, SE Antonarakis, G Hannon, MC Giddings, Y Ruan, B Wold, P Carninci, R Guigó, TR Gingeras (2012). "Landscape of transcription in human cells." *Nature* 489: 101-8.

J Harrow, A Frankish, JM Gonzalez, E Tapanari, M Diekhans, F Kokocinski, BL Aken, D Barrell, A Zadissa, S Searle, I Barnes, A Bignell, V Boychenko, T Hunt, M Kay, G Mukherjee, J Rajan, G

- Despacio-Reyes, G Saunders, C Steward, R Harte, M Lin, C Howald, A Tanzer, T Derrien, J Chrast, N Walters, S Balasubramanian, B Pei, M Tress, JM Rodriguez, I Ezkurdia, J van Baren, M Brent, D Haussler, M Kellis, A Valencia, A Reymond, M Gerstein, R Guigó, TJ Hubbard (2012). "GENCODE: the reference human genome annotation for The ENCODE Project." *Genome Res* 22: 1760-74.
- KY Yip, C Cheng, N Bhardwaj, JB Brown, J Leng, A Kundaje, J Rozowsky, E Birney, P Bickel, M Snyder, M Gerstein (2012). "Classification of human genomic regions based on experimentally determined binding sites of more than 100 transcription-related factors." *Genome Biol* 13: R48.
- C Cheng, R Alexander, R Min, J Leng, KY Yip, J Rozowsky, KK Yan, X Dong, S Djebali, Y Ruan, CA Davis, P Carninci, T Lassman, TR Gingeras, R Guigó, E Birney, Z Weng, M Snyder, M Gerstein (2012). "Understanding transcriptional regulation by integrative analysis of transcription factor binding data." *Genome Res* 22: 1658-67.
- SG Landt, GK Marinov, A Kundaje, P Kheradpour, F Pauli, S Batzoglou, BE Bernstein, P Bickel, JB Brown, P Cayting, Y Chen, G DeSalvo, C Epstein, KI Fisher-Aylor, G Euskirchen, M Gerstein, J Gertz, AJ Hartemink, MM Hoffman, VR Iyer, YL Jung, S Karmakar, M Kellis, PV Kharchenko, Q Li, T Liu, XS Liu, L Ma, A Milosavljevic, RM Myers, PJ Park, MJ Pazin, MD Perry, D Raha, TE Reddy, J Rozowsky, N Shores, A Sidow, M Slattery, JA Stamatoyannopoulos, MY Tolstorukov, KP White, S Xi, PJ Farnham, JD Lieb, BJ Wold, M Snyder (2012). "ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia." *Genome Res* 22: 1813-31.
- B Pei, C Sisu, A Frankish, C Howald, L Habegger, XJ Mu, R Harte, S Balasubramanian, A Tanzer, M Diekhans, A Reymond, TJ Hubbard, J Harrow, MB Gerstein (2012). "The GENCODE pseudogene resource." *Genome Biol* 13: R51.
- X Dong, MC Greven, A Kundaje, S Djebali, JB Brown, C Cheng, TR Gingeras, M Gerstein, R Guigó, E Birney, Z Weng (2012). "Modeling gene expression using chromatin features in various cellular contexts." *Genome Biol* 13: R53.
- JQ Wu, M Seay, VP Schulz, M Hariharan, D Tuck, J Lian, J Du, M Shi, Z Ye, M Gerstein, MP Snyder, S Weissman (2012). "Tcf7 is an important regulator of the switch of self-renewal and differentiation in a multipotential hematopoietic cell line." *PLoS Genet* 8: e1002565.
- HY Lam, C Pan, MJ Clark, P Lacroute, R Chen, R Haraksingh, M O'Huallachain, MB Gerstein, JM Kidd, CD Bustamante, M Snyder (2012). "Detecting and annotating genetic variations using the HugeSeq pipeline." *Nat Biotechnol* 30: 226-9.
- J Cotney, J Leng, S Oh, LE DeMare, SK Reilly, MB Gerstein, JP Noonan (2012). "Chromatin state signatures associated with tissue-specific gene expression and enhancer activity in the embryonic limb." *Genome Res* 22: 1069-80.
- M Vidal, DW Chan, M Gerstein, M Mann, GS Omenn, D Tagle, S Sechi, S Sechi, Workshop Participants (2012). "The human proteome - a scientific opportunity for transforming diagnostics, therapeutics, and healthcare." *Clin Proteomics* 9: 6.
- MJ Bamshad, JA Shendure, D Valle, A Hamosh, JR Lupski, RA Gibbs, E Boerwinkle, RP Lifton, M Gerstein, M Gunel, S Mane, DA Nickerson, DA Nickerson, Centers for Mendelian Genomics (2012). "The Centers for Mendelian Genomics: a new large-scale initiative to identify the genes underlying rare Mendelian conditions." *Am J Med Genet A* 158A: 1523-5.
- L Habegger, S Balasubramanian, DZ Chen, E Khurana, A Sboner, A Harman, J Rozowsky, D Clarke, M Snyder, M Gerstein (2012). "VAT: a computational framework to functionally annotate variants in personal genomes within a cloud-computing environment." *Bioinformatics* 28: 2267-9.
- G Zilberman-Schapira, J Chen, M Gerstein (2012). "On sports and genes." *Recent Pat DNA Gene Seq* 6: 180-8.
- D Clarke, N Bhardwaj, MB Gerstein (2012). "Novel insights through the integration of structural and functional genomics data with protein networks." *J Struct Biol* 179: 320-6.

- TA Gianoulis, MA Griffin, DJ Spakowicz, BF Dunican, CJ Alpha, A Sboner, AM Sismour, C Kodira, M Egholm, GM Church, MB Gerstein*, SA Strobel* (2012). "Genomic analysis of the hydrocarbon-producing, cellulolytic, endophytic fungus *Ascocoryne sarcoides*." *PLoS Genet* 8: e1002558.
- R Chen, GI Mias, J Li-Pook-Than, L Jiang, HY Lam, R Chen, E Miriami, KJ Karczewski, M Hariharan, FE Dewey, Y Cheng, MJ Clark, H Im, L Habegger, S Balasubramanian, M O'Huallachain, JT Dudley, S Hillenmeyer, R Haraksingh, D Sharon, G Euskirchen, P Lacroute, K Bettinger, AP Boyle, M Kasowski, F Grubert, S Seki, M Garcia, M Whirl-Carrillo, M Gallardo, MA Blasco, PL Greenberg, P Snyder, TE Klein, RB Altman, AJ Butte, EA Ashley, M Gerstein, KC Nadeau, H Tang, M Snyder (2012). "Personal omics profiling reveals dynamic molecular and medical phenotypes." *Cell* 148: 1293-307.
- DG MacArthur, S Balasubramanian, A Frankish, N Huang, J Morris, K Walter, L Jostins, L Habegger, JK Pickrell, SB Montgomery, CA Albers, ZD Zhang, DF Conrad, G Lunter, H Zheng, Q Ayub, MA DePristo, E Banks, M Hu, RE Handsaker, JA Rosenfeld, M Fromer, M Jin, XJ Mu, E Khurana, K Ye, M Kay, GI Saunders, MM Suner, T Hunt, IH Barnes, C Amid, DR Carvalho-Silva, AH Bignell, C Snow, B Yngvadottir, S Bumpstead, DN Cooper, Y Xue, IG Romero, IG Romero, 1000 Genomes Project Consortium, J Wang, Y Li, RA Gibbs, SA McCarroll, ET Dermitzakis, JK Pritchard, JC Barrett, J Harrow, ME Hurles, MB Gerstein*, C Tyler-Smith* (2012). "A systematic survey of loss-of-function variants in human protein-coding genes." *Science* 335: 823-8.
- G Li, X Ruan, RK Auerbach, KS Sandhu, M Zheng, P Wang, HM Poh, Y Goh, J Lim, J Zhang, HS Sim, SQ Peh, FH Mulawadi, CT Ong, YL Orlov, S Hong, Z Zhang, S Landt, D Raha, G Euskirchen, CL Wei, W Ge, H Wang, C Davis, KI Fisher-Aylor, A Mortazavi, M Gerstein, T Gingeras, B Wold, Y Sun, MJ Fullwood, E Cheung, E Liu, WK Sung, M Snyder, Y Ruan (2012). "Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation." *Cell* 148: 84-98.
- J Du, J Leng, L Habegger, A Sboner, D McDermott, M Gerstein (2012). "IQSeq: integrated isoform quantification analysis based on next-generation sequencing." *PLoS One* 7: e29175.

-- 2011 --

- N Bhardwaj, D Clarke, M Gerstein (2011). "Systematic control of protein interactions for systems biology." *Proc Natl Acad Sci U S A* 108: 20279-80.
- HY Lam, MJ Clark, R Chen, R Chen, G Natsoulis, M O'Huallachain, FE Dewey, L Habegger, EA Ashley, MB Gerstein, AJ Butte, HP Ji, M Snyder (2011). "Performance comparison of whole-genome sequencing platforms." *Nat Biotechnol* . 30: 78-82.
- D Greenbaum, A Sboner, XJ Mu, M Gerstein (2011). "Genomics and privacy: implications of the new reality of closed data for the field." *PLoS Comput Biol* 7: e1002278.
- RR Haraksingh, A Abyzov, M Gerstein, AE Urban, M Snyder (2011). "Genome-wide mapping of copy number variation in humans: comparative analysis of high resolution array platforms." *PLoS One* 6: e27859.
- C Cheng, KK Yan, W Hwang, J Qian, N Bhardwaj, J Rozowsky, ZJ Lu, W Niu, P Alves, M Kato, M Snyder, M Gerstein (2011). "Construction and analysis of an integrated regulatory network derived from high-throughput sequencing data." *PLoS Comput Biol* 7: e1002190.
- C Cheng, C Shou, KY Yip, MB Gerstein (2011). "Genome-wide analysis of chromatin features identifies histone modification sensitive and insensitive yeast transcription factors." *Genome Biol* 12: R111.
- D Greenbaum, M Gerstein (2011). "The role of cloud computing in managing the deluge of potentially private genetic data." *Am J Bioeth* 11: 39-41.
- C Cheng, R Min, M Gerstein (2011). "TIP: A probabilistic method for identifying transcription factor target genes from ChIP-seq binding profiles." *Bioinformatics* 27: 3221-7.

- SC Flores, MB Gerstein (2011). "Predicting protein ligand binding motions with the conformation explorer." *BMC Bioinformatics* 12: 417.
- KY Yip, L Utz, S Sitwell, X Hu, SS Sidhu, BE Turk, M Gerstein, PM Kim (2011). "Identification of specificity determining residues in peptide recognition domains using an information theoretic approach applied to large-scale binding maps." *BMC Biol* 9: 53.
- C Cheng, M Gerstein (2011). "Modeling the relative relationship of transcription factor binding and histone modifications to gene expression levels in mouse embryonic stem cells." *Nucleic Acids Res* 40: 553-68.
- MR Tanas, A Sboner, AM Oliveira, MR Erickson-Johnson, J Hespelt, PJ Hanwright, J Flanagan, Y Luo, K Fenwick, R Natrajan, C Mitsopoulos, M Zvelebil, BL Hoch, SW Weiss, M Debiec-Rychter, R Sciort, RB West, AJ Lazar, A Ashworth, JS Reis-Filho, CJ Lord, MB Gerstein, MA Rubin, BP Rubin (2011). "Identification of a disease-defining gene fusion in epithelioid hemangioendothelioma." *Sci Transl Med* 3: 98ra82.
- N Bhardwaj, A Abyzov, D Clarke, C Shou, MB Gerstein (2011). "Integration of protein motions with molecular networks reveals different mechanisms for permanent and transient interactions." *Protein Sci* 20: 1745-54.
- A Sboner, XJ Mu, D Greenbaum, RK Auerbach, MB Gerstein (2011). "The real cost of sequencing: higher than you think!" *Genome Biol* 12: 125.
- J Rozowsky, A Abyzov, J Wang, P Alves, D Raha, A Harmanci, J Leng, R Bjornson, Y Kong, N Kitabayashi, N Bhardwaj, M Rubin, M Snyder, M Gerstein (2011). "AlleleSeq: analysis of allele-specific expression and binding in a network framework." *Mol Syst Biol* 7: 522.
- ZD Zhang, J Du, H Lam, A Abyzov, AE Urban, M Snyder, M Gerstein (2011). "Identification of genomic indels and structural variations using split reads." *BMC Genomics* 12: 375.
- MB Clark, PP Amaral, FJ Schlesinger, ME Dinger, RJ Taft, JL Rinn, CP Ponting, PF Stadler, KV Morris, A Morillon, JS Rozowsky, MB Gerstein, C Wahlestedt, Y Hayashizaki, P Carninci, TR Gingeras, JS Mattick (2011). "The reality of pervasive transcription." *PLoS Biol* 9: e1000625; discussion e1001102.
- KK Yan, M Gerstein (2011). "The spread of scientific information: insights from the web usage statistics in PLoS article-level metrics." *PLoS One* 6: e19917.
- XJ Mu, ZJ Lu, Y Kong, HY Lam, MB Gerstein (2011). "Analysis of genomic variation in non-coding elements using population-scale sequencing data from the 1000 Genomes Project." *Nucleic Acids Res* 39: 7058-76.
- D Greenbaum, M Gerstein (2011). "Social considerations in research: consider them but don't use them." *Am J Bioeth* 11: 31-2.
- ENCODE Project Consortium, RM Myers, J Stamatoyannopoulos, M Snyder, I Dunham, RC Hardison, BE Bernstein, TR Gingeras, WJ Kent, E Birney, B Wold, GE Crawford (2011). "A user's guide to the encyclopedia of DNA elements (ENCODE)." *PLoS Biol* 9: e1001046.
- J Fasolo, A Sboner, MG Sun, H Yu, R Chen, D Sharon, PM Kim, M Gerstein, M Snyder (2011). "Diverse protein kinase interactions identified by protein microarrays reveal novel connections between cellular processes." *Genes Dev* 25: 767-78.
- TA Gianoulis, A Agarwal, M Snyder, MB Gerstein (2011). "The CRIT framework for identifying cross patterns in systems biology and application to chemogenomics." *Genome Biol* 12: R32.
- N Nègre, CD Brown, L Ma, CA Bristow, SW Miller, U Wagner, P Kheradpour, ML Eaton, P Loriaux, R Sealfon, Z Li, H Ishii, RF Spokony, J Chen, L Hwang, C Cheng, RP Auburn, MB Davis, M Domanus, PK Shah, CA Morrison, J Zieba, S Suchy, L Senderowicz, A Victorsen, NA Bild, AJ Grundstad, D Hanley, DM MacAlpine, M Mannervik, K Venken, H Bellen, R White, M Gerstein, S Russell, RL Grossman, B Ren, JW Posakony, M Kellis, KP White (2011). "A cis-regulatory map of the Drosophila genome." *Nature* 471: 527-31.

- GM Euskirchen, RK Auerbach, E Davidov, TA Gianoulis, G Zhong, J Rozowsky, N Bhardwaj, MB Gerstein, M Snyder (2011). "Diverse roles and interactions of the SWI/SNF chromatin remodeling complex revealed using global approaches." *PLoS Genet* 7: e1002008.
- D Lipman, P Flicek, S Salzberg, M Gerstein, R Knight (2011). "Closure of the NCBI SRA and implications for the long-term future of genomics data storage." *Genome Biol.* 2011;12 :402.
- J Jee, J Rozowsky, KY Yip, L Lochovsky, R Bjornson, G Zhong, Z Zhang, Y Fu, J Wang, Z Weng, M Gerstein (2011). "ACT: aggregation and correlation toolbox for analyses of genome tracks." *Bioinformatics* 27: 1152-4.
- A Karpikov, J Rozowsky, M Gerstein (2011). "Tiling array data analysis: a multiscale approach using wavelets." *BMC Bioinformatics* 12: 57.
- A Abyzov, AE Urban, M Snyder, M Gerstein (2011). "CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing." *Genome Res* 21: 974-84.
- C Cheng, KK Yan, KY Yip, J Rozowsky, R Alexander, C Shou, M Gerstein (2011). "A statistical framework for modeling gene expression using chromatin features and application to modENCODE datasets." *Genome Biol* 12: R15.
- MF Berger, MS Lawrence, F Demichelis, Y Drier, K Cibulskis, AY Sivachenko, A Sboner, R Esgueva, D Pflueger, C Sougnez, R Onofrio, SL Carter, K Park, L Habegger, L Ambrogio, T Fennell, M Parkin, G Saksena, D Voet, AH Ramos, TJ Pugh, J Wilkinson, S Fisher, W Winckler, S Mahan, K Ardlie, J Baldwin, JW Simons, N Kitabayashi, TY MacDonald, PW Kantoff, L Chin, SB Gabriel, MB Gerstein, TR Golub, M Meyerson, A Tewari, ES Lander, G Getz, MA Rubin, LA Garraway (2011). "The genomic complexity of primary human prostate cancer." *Nature* 470: 214-20.
- RE Mills, K Walter, C Stewart, RE Handsaker, K Chen, C Alkan, A Abyzov, SC Yoon, K Ye, RK Cheetham, A Chinwalla, DF Conrad, Y Fu, F Grubert, I Hajirasouliha, F Hormozdiari, LM Iakoucheva, Z Iqbal, S Kang, JM Kidd, MK Konkel, J Korn, E Khurana, D Kural, HY Lam, J Leng, R Li, Y Li, CY Lin, R Luo, XJ Mu, J Nemesh, HE Peckham, T Rausch, A Scally, X Shi, MP Stromberg, AM Stütz, AE Urban, JA Walker, J Wu, Y Zhang, ZD Zhang, MA Batzer, L Ding, GT Marth, G McVean, J Sebat, M Snyder, J Wang, K Ye, EE Eichler*, MB Gerstein*, ME Hurles*, C Lee*, SA McCarroll*, JO Korbel*, 1000 Genomes Project (2011). "Mapping copy number variation by population-scale genome sequencing." *Nature* 470: 59-65.
- C Shou, N Bhardwaj, HY Lam, KK Yan, PM Kim, M Snyder, MB Gerstein (2011). "Measuring the evolutionary rewiring of biological networks." *PLoS Comput Biol* 7: e1001050.
- A Abyzov, M Gerstein (2011). "AGE: defining breakpoints of genomic structural variants at single-nucleotide resolution, through optimal alignments with gap excision." *Bioinformatics* 27: 595-603.
- S Balasubramanian, L Habegger, A Frankish, DG MacArthur, R Harte, C Tyler-Smith, J Harrow, M Gerstein (2011). "Gene inactivation and its implications for annotation in the era of personal genomics." *Genes Dev* 25: 1-10.
- FM Vaccarino, AE Urban, HE Stevens, A Szekely, A Abyzov, EL Grigorenko, M Gerstein, S Weissman (2011). "Annual Research Review: The promise of stem cell research for neuropsychiatric disorders." *J Child Psychol Psychiatry* 52: 504-16.
- ZJ Lu, KY Yip, G Wang, C Shou, LW Hillier, E Khurana, A Agarwal, R Auerbach, J Rozowsky, C Cheng, M Kato, DM Miller, F Slack, M Snyder, RH Waterston, V Reinke, MB Gerstein (2011). "Prediction and characterization of noncoding RNAs in *C. elegans* by integrating conservation, secondary structure, and high-throughput sequencing and array data." *Genome Res* 21: 276-85.
- W Niu, ZJ Lu, M Zhong, M Sarov, JI Murray, CM Brdlik, J Janette, C Chen, P Alves, E Preston, C Slightam, L Jiang, AA Hyman, SK Kim, RH Waterston, M Gerstein, M Snyder, V Reinke (2011). "Diverse transcription factor binding features revealed by genome-wide ChIP-seq in *C. elegans*." *Genome Res* 21: 245-54.

- L Habegger, A Sboner, TA Gianoulis, J Rozowsky, A Agarwal, M Snyder, M Gerstein (2011). "RSEQtools: a modular framework to analyze RNA-Seq data using compact, anonymized data summaries." *Bioinformatics* 27: 281-3.
- D Pflueger, S Terry, A Sboner, L Habegger, R Esgueva, PC Lin, MA Svensson, N Kitabayashi, BJ Moss, TY MacDonald, X Cao, T Barrette, AK Tewari, MS Chee, AM Chinnaiyan, DS Rickman, F Demichelis, MB Gerstein, MA Rubin (2011). "Discovery of non-ETS gene fusions in human prostate cancer using next-generation RNA sequencing." *Genome Res* 21: 56-67.

-- 2010 --

- Reproducible Research: Addressing the need for data and code sharing in computational science. Yale Law School Roundtable on Data and Code Sharing (2010) *Computing in Science & Engineering* 12(5): 8-13 (Sept/Oct).
- MB Gerstein, ZJ Lu, EL Van Nostrand, C Cheng, BI Arshinoff, T Liu, KY Yip, R Robilotto, A Rechtsteiner, K Ikegami, P Alves, A Chateigner, M Perry, M Morris, RK Auerbach, X Feng, J Leng, A Vielle, W Niu, K Rhrissorakrai, A Agarwal, RP Alexander, G Barber, CM Brdlik, J Brennan, JJ Brouillet, A Carr, MS Cheung, H Clawson, S Contrino, LO Dannenberg, AF Dernburg, A Desai, L Dick, AC Dosé, J Du, T Egelhofer, S Ercan, G Euskirchen, B Ewing, EA Feingold, R Gassmann, PJ Good, P Green, F Gullier, M Gutwein, MS Guyer, L Habegger, T Han, JG Henikoff, SR Henz, A Hinrichs, H Holster, T Hyman, AL Iniguez, J Janette, M Jensen, M Kato, WJ Kent, E Kephart, V Khivansara, E Khurana, JK Kim, P Kolasinska-Zwierz, EC Lai, I Latorre, A Leahey, S Lewis, P Lloyd, L Lochovsky, RF Lowdon, Y Lubling, R Lyne, M MacCoss, SD Mackowiak, M Mangone, S McKay, D Mecnas, G Merrihew, DM Miller, A Muroyama, JI Murray, SL Ooi, H Pham, T Phippen, EA Preston, N Rajewsky, G Rättsch, H Rosenbaum, J Rozowsky, K Rutherford, P Ruzanov, M Sarov, R Sasidharan, A Sboner, P Scheid, E Segal, H Shin, C Shou, FJ Slack, C Slightam, R Smith, WC Spencer, EO Stinson, S Taing, T Takasaki, D Vafeados, K Voronina, G Wang, NL Washington, CM Whittle, B Wu, KK Yan, G Zeller, Z Zha, M Zhong, X Zhou, modENCODE Consortium, J Ahringer, S Strome, KC Gunsalus, G Micklem, XS Liu, V Reinke, SK Kim, LW Hillier, S Henikoff, F Piano, M Snyder, L Stein, JD Lieb, RH Waterston (2010). "Integrative analysis of the *Caenorhabditis elegans* genome by the modENCODE project." *Science* 330: 1775-87.
- N Bhardwaj, PM Kim, MB Gerstein (2010). "Rewiring of transcriptional regulatory networks: hierarchy, rather than connectivity, better reflects the importance of regulators." *Sci Signal* 3: ra79.
- X Li, TA Gianoulis, KY Yip, M Gerstein, M Snyder (2010). "Extensive in vivo metabolite-protein interactions revealed by large-scale systematic analyses." *Cell* 143: 639-50.
- ZD Zhang, MB Gerstein (2010). "Detection of copy number variation from array intensity and sequencing read depth using a stepwise Bayesian model." *BMC Bioinformatics* 11: 539.
- 1000 Genomes Project Consortium, RM Durbin, GR Abecasis, DL Altshuler, A Auton, LD Brooks, RM Durbin, RA Gibbs, ME Hurles, GA McVean (2010). "A map of human genome variation from population-scale sequencing." *Nature* 467: 1061-73.
- A Sboner, L Habegger, D Pflueger, S Terry, DZ Chen, JS Rozowsky, AK Tewari, N Kitabayashi, BJ Moss, MS Chee, F Demichelis, MA Rubin, MB Gerstein (2010). "FusionSeq: a modular framework for finding gene fusions by analyzing paired-end RNA-sequencing data." *Genome Biol* 11: R104.
- KH Cheung, M Samwald, RK Auerbach, MB Gerstein (2010). "Structured digital tables on the Semantic Web: toward a structured digital literature." *Mol Syst Biol* 6: 403.
- RP Alexander, G Fang, J Rozowsky, M Snyder, MB Gerstein (2010). "Annotating non-coding regions of the genome." *Nat Rev Genet* 11: 559-71.
- E Khurana, HY Lam, C Cheng, N Carriero, P Cayting, MB Gerstein (2010). "Segmental duplications in the human genome reveal details of pseudogene formation." *Nucleic Acids Res* 38: 6997-7007.

- A Agarwal, D Koppstein, J Rozowsky, A Sboner, L Habegger, LW Hillier, R Sasidharan, V Reinke, RH Waterston, M Gerstein (2010). "Comparison and calibration of transcriptome data from RNA-Seq and tiling arrays." *BMC Genomics* 11: 383.
- ME Holford, E Khurana, KH Cheung, M Gerstein (2010). "Using semantic web rules to reason on an ontology of pseudogenes." *Bioinformatics* 26: i71-8.
- N Bhardwaj, MB Carson, A Abyzov, KK Yan, H Lu, MB Gerstein (2010). "Analysis of combinatorial regulation: scaling of partnerships between regulators with the number of governed targets." *PLoS Comput Biol* 6: e1000755.
- NR Voss, M Gerstein (2010). "3V: cavity, channel and cleft volume calculator and extractor." *Nucleic Acids Res* 38: W555-62.
- HY Lam, PM Kim, J Mok, R Tonikian, SS Sidhu, BE Turk, M Snyder, MB Gerstein (2010). "MOTIPS: automated motif analysis for predicting targets of modular protein domains." *BMC Bioinformatics* 11: 243.
- KK Yan, G Fang, N Bhardwaj, RP Alexander, M Gerstein (2010). "Comparing genomes to computer operating systems in terms of the topology and evolution of their regulatory control networks." *Proc Natl Acad Sci U S A* 107: 9186-91.
- PV Patel, TA Gianoulis, RD Bjornson, KY Yip, DM Engelman, MB Gerstein (2010). "Analysis of membrane proteins in metagenomics: networks of correlated environmental features and protein families." *Genome Res* 20: 960-71.
- Y Lee, X Yang, Y Huang, H Fan, Q Zhang, Y Wu, J Li, R Hasina, C Cheng, MW Lingen, MB Gerstein, RR Weichselbaum, HR Xing, YA Lussier (2010). "Network modeling identifies molecular functions targeted by miR-204 to suppress head and neck tumor metastasis." *PLoS Comput Biol* 6: e1000730.
- G Fang, N Bhardwaj, R Robilotto, MB Gerstein (2010). "Getting started in gene orthology and functional analysis." *PLoS Comput Biol* 6: e1000703.
- N Bhardwaj, KK Yan, MB Gerstein (2010). "Analysis of diverse regulatory networks in a hierarchical context shows consistent tendencies for collaboration in the middle levels." *Proc Natl Acad Sci U S A* 107: 6841-6.
- M Kasowski, F Grubert, C Heffelfinger, M Hariharan, A Asabere, SM Waszak, L Habegger, J Rozowsky, M Shi, AE Urban, MY Hong, KJ Karczewski, W Huber, SM Weissman, MB Gerstein, JO Korbel, M Snyder (2010). "Variation in transcription factor binding among humans." *Science* 328: 232-5.
- A Sboner, F Demichelis, S Calza, Y Pawitan, SR Setlur, Y Hoshida, S Perner, HO Adami, K Fall, LA Mucci, PW Kantoff, M Stampfer, SO Andersson, E Varenhorst, JE Johansson, MB Gerstein, TR Golub, MA Rubin, O Andr n (2010). "Molecular sampling of prostate cancer: a dilemma for predicting disease progression." *BMC Med Genomics* 3: 8.
- ZD Zhang, A Frankish, T Hunt, J Harrow, M Gerstein (2010). "Identification and analysis of unitary pseudogenes: historic and contemporary gene losses in humans and other primates." *Genome Biol* 11: R26.
- N Bhardwaj, M Gerstein, H Lu (2010). "Genome-wide sequence-based prediction of peripheral proteins using a novel semi-supervised learning technique." *BMC Bioinformatics* 11 Suppl 1: S6.
- JQ Wu, L Habegger, P Noisa, A Szekely, C Qiu, S Hutchison, D Raha, M Egholm, H Lin, S Weissman, W Cui, M Gerstein, M Snyder (2010). "Dynamic transcriptomes during neural differentiation of human embryonic stem cells revealed by short, long, and paired-end sequencing." *Proc Natl Acad Sci U S A* 107: 5254-9.
- M Snyder, J Du, M Gerstein (2010). "Personal genome sequencing: current approaches and challenges." *Genes Dev* 24: 423-31.
- M Zhong, W Niu, ZJ Lu, M Sarov, JI Murray, J Janette, D Raha, KL Sheaffer, HY Lam, E Preston, C Slightam, LW Hillier, T Brock, A Agarwal, R Auerbach, AA Hyman, M Gerstein, SE Mango, SK

Kim, RH Waterston, V Reinke, M Snyder (2010). "Genome-wide identification of binding sites defines distinct functions for *Caenorhabditis elegans* PHA-4/FOXA in development and environmental response." *PLoS Genet* 6: e1000848.

J Mok, PM Kim, HY Lam, S Piccirillo, X Zhou, GR Jeschke, DL Sheridan, SA Parker, V Desai, M Jwa, E Camerini, H Niu, M Good, A Remenyi, JL Ma, YJ Sheu, HE Sassi, R Sopko, CS Chan, C De Virgilio, NM Hollingsworth, WA Lim, DF Stern, B Stillman, BJ Andrews, MB Gerstein, M Snyder, BE Turk (2010). "Deciphering protein kinase specificity through large-scale analysis of yeast phosphorylation site motifs." *Sci Signal* 3: ra12.

D Raha, Z Wang, Z Moqtaderi, L Wu, G Zhong, M Gerstein, K Struhl, M Snyder (2010). "Close association of RNA polymerase II and many transcription factors with Pol III genes." *Proc Natl Acad Sci U S A* 107: 3639-44.

KY Yip, RP Alexander, KK Yan, M Gerstein (2010). "Improved reconstruction of in silico gene regulatory networks by integrating knockout and perturbation data." *PLoS One* 5: e8121.

HY Lam, XJ Mu, AM Stütz, A Tanzer, PD Cayting, M Snyder, PM Kim, JO Korb, MB Gerstein (2010). "Nucleotide-resolution analysis of structural variants using BreakSeq and a breakpoint library." *Nat Biotechnol* 28: 47-55.

A Abyzov, R Bjornson, M Felipe, M Gerstein (2010). "RigidFinder: a fast and sensitive method to detect rigid blocks in large macromolecular complexes." *Proteins* 78: 309-24.

-- 2009 --

Science and the Law: Grappling with the Gulf. D Greenbaum, MB Gerstein (2009) *Science* 323: 210.

A Canaan, I Haviv, AE Urban, VP Schulz, S Hartman, Z Zhang, D Palejev, AB Deisseroth, J Lacy, M Snyder, M Gerstein, SM Weissman (2009). "EBNA1 regulates cellular gene expression by binding cellular promoters." *Proc Natl Acad Sci U S A* 106: 22421-6.

D Greenbaum, M Gerstein (2009). "Social networking and personal genomics: suggestions for optimizing the interaction." *Am J Bioeth* 9: 15-9.

R Tonikian, X Xin, CP Toret, D Gfeller, C Landgraf, S Panni, S Paoluzi, L Castagnoli, B Currell, S Seshagiri, H Yu, B Winsor, M Vidal, MB Gerstein, GD Bader, R Volkmer, G Cesareni, DG Drubin, PM Kim, SS Sidhu, C Boone (2009). "Bayesian modeling of the yeast SH3 domain interactome predicts spatiotemporal dynamics of endocytosis proteins." *PLoS Biol* 7: e1000218.

YJ Liu, D Zheng, S Balasubramanian, N Carriero, E Khurana, R Robilotto, MB Gerstein (2009). "Comprehensive analysis of the pseudogenes of glycolytic enzymes in vertebrates: the anomalously high number of GAPDH pseudogenes highlights a recent burst of retrotrans-positional activity." *BMC Genomics* 10: 480.

A Sboner, A Karpikov, G Chen, M Smith, D Mattoon, M Dawn, L Freeman-Cook, B Schweitzer, MB Gerstein (2009). "Robust-linear-model normalization to reduce technical variability in functional protein microarrays." *J Proteome Res* 8: 5451-64.

C Cheng, N Bhardwaj, M Gerstein (2009). "The relationship between the evolution of microRNA targets and the length of their UTRs." *BMC Genomics* 10: 431.

Y Arinaminpathy, E Khurana, DM Engelman, MB Gerstein (2009). "Computational analysis of membrane proteins: the largest class of drug targets." *Drug Discov Today* 14: 1130-5.

C Cheng, X Fu, P Alves, M Gerstein (2009). "mRNA expression profiles show differential regulatory effects of microRNAs between estrogen receptor-positive and estrogen receptor-negative breast cancer." *Genome Biol* 10: R90.

RK Auerbach, G Euskirchen, J Rozowsky, N Lamarre-Vincent, Z Moqtaderi, P Lefrançois, K Struhl, M Gerstein, M Snyder (2009). "Mapping accessible chromatin regions using Sono-Seq." *Proc Natl Acad Sci U S A* 106: 14926-31.

- KY Yip, PM Kim, D McDermott, M Gerstein (2009). "Multi-level learning: improving the prediction of protein, domain and residue interactions by allowing information flow between levels." *BMC Bioinformatics* 10: 241.
- A Rzhetsky, M Seringhaus, MB Gerstein (2009). "Getting started in text mining: part two." *PLoS Comput Biol* 5: e1000411.
- D Pflueger, DS Rickman, A Sboner, S Perner, CJ LaFargue, MA Svensson, BJ Moss, N Kitabayashi, Y Pan, A de la Taille, R Kuefer, AK Tewari, F Demichelis, MS Chee, MB Gerstein, MA Rubin (2009). "N-myc downstream regulated gene 1 (NDRG1) is fused to ERG in prostate cancer." *Neoplasia* 11: 804-11.
- X Guo, Z Zhang, MB Gerstein, D Zheng (2009). "Small RNAs originated from pseudogenes: cis- or trans-acting?" *PLoS Comput Biol* 5: e1000449.
- RP Alexander, PM Kim, T Emonet, MB Gerstein (2009). "Understanding modularity in molecular networks requires dynamics." *Sci Signal* 2: pe44.
- R Sasidharan, A Agarwal, J Rozowsky, M Gerstein (2009). "An approach to comparing tiling array and high throughput sequencing technologies for genomic transcript mapping." *BMC Res Notes* 2: 150.
- K Talbert-Slagle, S Marlatt, FN Barrera, E Khurana, J Oates, M Gerstein, DM Engelman, AM Dixon, D Dimairo (2009). "Artificial transmembrane oncoproteins smaller than the bovine papillomavirus E5 protein redefine sequence requirements for activation of the platelet-derived growth factor beta receptor." *J Virol* 83: 9773-85.
- JO Korbelt, T Tirosh-Wagner, AE Urban, XN Chen, M Kasowski, L Dai, F Grubert, C Erdman, MC Gao, K Lange, EM Sobel, GM Barlow, AS Aylsworth, NJ Carpenter, RD Clark, MY Cohen, E Doran, T Falik-Zaccai, SO Lewin, IT Lott, BC McGillivray, JB Moeschler, MJ Pettenati, SM Puschel, KW Rao, LG Shaffer, M Shohat, AJ Van Riper, D Warburton, S Weissman, MB Gerstein, M Snyder, JR Korenberg (2009). "The genetic architecture of Down syndrome phenotypes revealed by high-resolution analysis of human segmental trisomies." *Proc Natl Acad Sci U S A* 106: 12031-6.
- J Du, RD Bjornson, ZD Zhang, Y Kong, M Snyder, MB Gerstein (2009). "Integrating sequencing technologies in personal genomics: optimal low cost reconstruction of structural variants." *PLoS Comput Biol* 5: e1000432.
- SE Celniker, LA Dillon, MB Gerstein, KC Gunsalus, S Henikoff, GH Karpen, M Kellis, EC Lai, JD Lieb, DM MacAlpine, G Micklem, F Piano, M Snyder, L Stein, KP White, RH Waterston, modENCODE Consortium (2009). "Unlocking the secrets of the genome." *Nature* 459: 927-30.
- Y Xia, EA Franzosa, MB Gerstein (2009). "Integrated assessment of genomic correlates of protein evolutionary rate." *PLoS Comput Biol* 5: e1000413.
- L Ni, C Bruce, C Hart, J Leigh-Bell, D Gelperin, L Umansky, MB Gerstein, M Snyder (2009). "Dynamic and complex transcription factor binding during an inducible response in yeast." *Genes Dev* 23: 1351-63.
- N Bhardwaj, M Gerstein (2009). "Relating protein conformational changes to packing efficiency and disorder." *Protein Sci* 18: 1230-40.
- M Snyder, S Weissman, M Gerstein (2009). "Personal phenotypes to go with personal genomes." *Mol Syst Biol* 5: 273.
- C Cheng, LM Li, P Alves, M Gerstein (2009). "Systematic identification of transcription factors associated with patient survival in cancers." *BMC Genomics* 10: 225.
- Y Mishima, C Abreu-Goodger, AA Staton, C Stahlhut, C Shou, C Cheng, M Gerstein, AJ Enright, AJ Giraldez (2009). "Zebrafish miR-1 and miR-133 shape muscle gene expression and regulate sarcomeric actin organization." *Genes Dev* 23: 619-32.

- JO Korbelt, A Abyzov, XJ Mu, N Carriero, P Cayting, Z Zhang, M Snyder, MB Gerstein (2009). "PEMer: a computational framework with simulation-based error models for inferring genomic structural variants from massive paired-end sequencing data." *Genome Biol* 10: R23.
- KS Keating, SC Flores, MB Gerstein, LA Kuhn (2009). "StoneHinge: hinge prediction by network analysis of individual protein structures." *Protein Sci* 18: 359-71.
- TA Gianoulis, J Raes, PV Patel, R Bjornson, JO Korbelt, I Letunic, T Yamada, A Paccanaro, LJ Jensen, M Snyder, P Bork, MB Gerstein (2009). "Quantifying environmental adaptation of metabolic pathways in metagenomics." *Proc Natl Acad Sci U S A* 106: 1374-9.
- P Lefrançois, GM Euskirchen, RK Auerbach, J Rozowsky, T Gibson, CM Yellman, M Gerstein, M Snyder (2009). "Efficient yeast ChIP-Seq using multiplex short-read DNA sequencing." *BMC Genomics* 10: 37.
- F Demichelis, SR Setlur, R Beroukhim, S Perner, JO Korbelt, CJ Lafargue, D Pflueger, C Pina, MD Hofer, A Sboner, MA Svensson, DS Rickman, A Urban, M Snyder, M Meyerson, C Lee, MB Gerstein, R Kuefer, MA Rubin (2009). "Distinct genomic aberrations associated with ERG rearranged prostate cancer." *Genes Chromosomes Cancer* 48: 366-80.
- X Zhang, Z Lian, C Padden, MB Gerstein, J Rozowsky, M Snyder, TR Gingeras, P Kapranov, SM Weissman, PE Newburger (2009). "A myelopoiesis-associated regulatory intergenic noncoding RNA transcript within the human HOXA cluster." *Blood* 113: 2526-34.
- S Balasubramanian, D Zheng, YJ Liu, G Fang, A Frankish, N Carriero, R Robilotto, P Cayting, M Gerstein (2009). "Comparative analysis of processed ribosomal protein pseudogenes in four mammalian genomes." *Genome Biol* 10: R2.
- J Rozowsky, G Euskirchen, RK Auerbach, ZD Zhang, T Gibson, R Bjornson, N Carriero, M Snyder, MB Gerstein (2009). "PeakSeq enables systematic scoring of ChIP-seq experiments relative to controls." *Nat Biotechnol* 27: 66-75.
- SC Popescu, GV Popescu, S Bachan, Z Zhang, M Gerstein, M Snyder, SP Dinesh-Kumar (2009). "MAPK target networks in Arabidopsis thaliana revealed using functional protein microarrays." *Genes Dev* 23: 80-92.
- LY Wang, A Abyzov, JO Korbelt, M Snyder, M Gerstein (2009). "MSB: a mean-shift-based approach for the analysis of structural variation in the genome." *Genome Res* 19: 106-17.
- Z Wang, M Gerstein, M Snyder (2009). "RNA-Seq: a revolutionary tool for transcriptomics." *Nat Rev Genet* 10: 57-63.
- KY Yip, M Gerstein (2009). "Training set expansion: an approach to improving the reconstruction of biological networks from limited and uneven reliable interactions." *Bioinformatics* 25: 243-50.
- HY Lam, E Khurana, G Fang, P Cayting, N Carriero, KH Cheung, MB Gerstein (2009). "Pseudofam: the pseudogene families database." *Nucleic Acids Res* 37: D738-43.

-- 2008 --

- Genomics Confounds Gene Classification. M Seringhaus, M Gerstein (2008) *American Scientist* 96:466-473 (Nov-Dec).
- M Seringhaus, J Rozowsky, T Royce, U Nagalakshmi, J Jee, M Snyder, M Gerstein (2008). "Mismatch oligonucleotides in human and yeast: guidelines for probe design on tiling microarrays." *BMC Genomics* 9: 635.
- AY Yam, Y Xia, HT Lin, A Burlingame, M Gerstein, J Frydman (2008). "Defining the TRiC/CCT interactome links chaperonin function to stabilization of newly made proteins with complex topologies." *Nat Struct Mol Biol* 15: 1255-62.
- D Greenbaum, J Du, M Gerstein (2008). "Genomic anonymity: have we already lost it?" *Am J Bioeth* 8: 71-4.

- Y Hasin, T Olender, M Khen, C Gonzaga-Jauregui, PM Kim, AE Urban, M Snyder, MB Gerstein, D Lancet, JO Korbelt (2008). "High-resolution copy-number variation map reflects human olfactory receptor diversity and evolution." *PLoS Genet* 4: e1000249.
- PM Kim, HY Lam, AE Urban, JO Korbelt, J Affourtit, F Grubert, X Chen, S Weissman, M Snyder, MB Gerstein (2008). "Analysis of copy number variants and segmental duplications in the human genome: Evidence for a change in the process of formation in recent evolutionary history." *Genome Res* 18: 1865-74.
- ZD Zhang, J Rozowsky, M Snyder, J Chang, M Gerstein (2008). "Modeling ChIP sequencing in silico with applications." *PLoS Comput Biol* 4: e1000158.
- R Sasidharan, A Smith, M Gerstein (2008). "Transmembrane protein oxygen content and compartmentalization of cells." *PLoS One* 3: e2726.
- A Rzhetsky, M Seringhaus, M Gerstein (2008). "Seeking a new biology through text mining." *Cell* 134: 9-13.
- R Sasidharan, M Gerstein (2008). "Genomics: protein fossils live on as RNA." *Nature* 453: 729-31.
- JO Korbelt, PM Kim, X Chen, AE Urban, S Weissman, M Snyder, MB Gerstein (2008). "The current excitement about copy-number variation: how it relates to gene duplications and protein families." *Curr Opin Struct Biol* 18: 366-74.
- YJ Huang, D Hang, LJ Lu, L Tong, MB Gerstein, GT Montelione (2008). "Targeting the human cancer pathway protein interaction network by structural genomics." *Mol Cell Proteomics* 7: 2048-60.
- Z Lian, A Karpikov, J Lian, MC Mahajan, S Hartman, M Gerstein, M Snyder, SM Weissman (2008). "A genomic analysis of RNA polymerase II modification and chromatin architecture related to 3' end RNA polyadenylation." *Genome Res* 18: 1224-37.
- KD Mertz, F Demichelis, A Sboner, MS Hirsch, P Dal Cin, K Struckmann, M Storz, S Scherrer, DM Schmid, RT Strebel, NM Probst-Hensch, M Gerstein, H Moch, MA Rubin (2008). "Association of cytokeratin 7 and 19 expression with genomic stability and favorable prognosis in clear cell renal cell cancer." *Int J Cancer* 123: 569-76.
- U Nagalakshmi, Z Wang, K Waern, C Shou, D Raha, M Gerstein, M Snyder (2008). "The transcriptional landscape of the yeast genome defined by RNA sequencing." *Science* 320: 1344-9.
- SC Flores, KS Keating, J Painter, F Morcos, K Nguyen, EA Merritt, LA Kuhn, MB Gerstein (2008). "HingeMaster: normal mode hinge prediction approach and integration of complementary predictors." *Proteins* 73: 299-319.
- ZD Zhang, G Weinstock, M Gerstein (2008). "Rapid evolution by positive Darwinian selection in T-cell antigen CD4 in primates." *J Mol Evol* 66: 446-56.
- PE Bourne, JL Fink, M Gerstein (2008). "Open access: taking full advantage of the content." *PLoS Comput Biol* 4: e1000037.
- PM Kim, A Sboner, Y Xia, M Gerstein (2008). "The role of disorder in interaction networks: a structural analysis." *Mol Syst Biol* 4: 179.
- M Seringhaus, M Gerstein (2008). "Manually structured digital abstracts: a scaffold for automatic text mining." *FEBS Lett* 582: 1170.
- DS Johnson, W Li, DB Gordon, A Bhattacharjee, B Curry, J Ghosh, L Brizuela, JS Carroll, M Brown, P Flicek, CM Koch, I Dunham, M Bieda, X Xu, PJ Farnham, P Kapranov, DA Nix, TR Gingeras, X Zhang, H Holster, N Jiang, RD Green, JS Song, SA McCuine, E Anton, L Nguyen, ND Trinklein, Z Ye, K Ching, D Hawkins, B Ren, PC Scacheri, J Rozowsky, A Karpikov, G Euskirchen, S Weissman, M Gerstein, M Snyder, A Yang, Z Moqtaderi, H Hirsch, HP Shulha, Y Fu, Z Weng, K Struhl, RM Myers, JD Lieb, XS Liu (2008). "Systematic evaluation of variability in ChIP-chip experiments using predefined DNA targets." *Genome Res* 18: 393-403.

- MR Seringhaus, PD Cayting, MB Gerstein (2008). "Uncovering trends in gene naming." *Genome Biol* 9: 401.
- JQ Wu, J Du, J Rozowsky, Z Zhang, AE Urban, G Euskirchen, S Weissman, M Gerstein, M Snyder (2008). "Systematic analysis of transcribed loci in ENCODE regions using RACE sequencing reveals extensive transcription in the human genome." *Genome Biol* 9: R3.
- ZD Zhang, P Cayting, G Weinstock, M Gerstein (2008). "Analysis of nuclear receptor pseudogenes in vertebrates: how the silent tell their stories." *Mol Biol Evol* 25: 131-43.
- KY Yip, P Patel, PM Kim, DM Engelman, D McDermott, M Gerstein (2008). "An integrated system for studying residue coevolution in proteins." *Bioinformatics* 24: 290-2.

-- 2007 --

- Semantic Web Approach to Database Integration in the Life Sciences. KH Cheung, AK Smith, KYL Yip, CJO Baker, MB Gerstein (2007) in *Semantic Web: Revolutionizing Knowledge Discovery in the Life Sciences* (eds. C Baker and K Cheung, Springer, NY), pp. 11-30. *Semantic Web Standards: Legal and Social Issues and Implications*. D Greenbaum, M Gerstein (2007) in *Semantic Web: Revolutionizing Knowledge Discovery in the Life Sciences* (eds. C Baker and K Cheung, Springer, NY), pp. 413-433.
- PM Kim, JO Korbelt, MB Gerstein (2007). "Positive selection at the protein network periphery: evaluation in terms of structural constraints and cellular context." *Proc Natl Acad Sci U S A* 104: 20274-9.
- SR Setlur, TE Royce, A Sboner, JM Mosquera, F Demichelis, MD Hofer, KD Mertz, M Gerstein, MA Rubin (2007). "Integrative microarray analysis of pathways dysregulated in metastatic prostate cancer." *Cancer Res* 67: 10296-303.
- A Smith, K Cheung, M Krauthammer, M Schultz, M Gerstein (2007). "Leveraging the structure of the Semantic Web to enhance information retrieval for proteomics." *Bioinformatics* 23: 3073-9.
- AJ McClellan, Y Xia, AM Deutschbauer, RW Davis, M Gerstein, J Frydman (2007). "Diverse cellular functions of the Hsp90 molecular chaperone uncovered using systems approaches." *Cell* 131: 121-35.
- JO Korbelt, AE Urban, JP Affourtit, B Godwin, F Grubert, JF Simons, PM Kim, D Palejev, NJ Carriero, L Du, BE Taillon, Z Chen, A Tanzer, AC Saunders, J Chi, F Yang, NP Carter, ME Hurles, SM Weissman, TT Harkins, MB Gerstein, M Egholm, M Snyder (2007). "Paired-end mapping reveals extensive structural variation in the human genome." *Science* 318: 420-6.
- EZ Yu, AE Burba, M Gerstein (2007). "PARE: a tool for comparing protein abundance and mRNA expression data." *BMC Bioinformatics* 8: 309.
- AR Borneman, TA Gianoulis, ZD Zhang, H Yu, J Rozowsky, MR Seringhaus, LY Wang, M Gerstein, M Snyder (2007). "Divergence of transcription factor binding sites across related yeast species." *Science* 317: 815-9.
- S Orchard, L Salwinski, S Kerrien, L Montecchi-Palazzi, M Oesterheld, V Stümpflen, A Ceol, A Chatr-aryamontri, J Armstrong, P Woollard, JJ Salama, S Moore, J Wojcik, GD Bader, M Vidal, ME Cusick, M Gerstein, AC Gavin, G Superti-Furga, J Greenblatt, J Bader, P Uetz, M Tyers, P Legrain, S Fields, N Mulder, M Gilson, M Niepmann, L Burgoon, J De Las Rivas, C Prieto, VM Perreau, C Hogue, HW Mewes, R Apweiler, I Xenarios, D Eisenberg, G Cesareni, H Hermjakob (2007). "The minimum information required for reporting a molecular interaction experiment (MIMIx)." *Nat Biotechnol* 25: 894-8.
- TE Royce, JS Rozowsky, MB Gerstein (2007). "Toward a universal microarray: prediction of gene expression through nearest-neighbor probe sequence identification." *Nucleic Acids Res* 35: e99.
- AR Borneman, ZD Zhang, J Rozowsky, MR Seringhaus, M Gerstein, M Snyder (2007). "Transcription factor binding site identification in yeast: a comparison of high-density oligonucleotide and PCR-based microarray platforms." *Funct Integr Genomics* 7: 335-45.

- SC Flores, MB Gerstein (2007). "FlexOracle: predicting flexible hinges by identification of stable domains." *BMC Bioinformatics* 8: 215.
- LJ Lu, A Sboner, YJ Huang, HX Lu, TA Gianoulis, KY Yip, PM Kim, GT Montelione, MB Gerstein (2007). "Comparing classical pathways and modern networks: towards the development of an edge ontology." *Trends Biochem Sci* 32: 320-31.
- ENCODE Project Consortium (2007). "Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project." *Nature* 447: 799-816.
- GM Euskirchen, JS Rozowsky, CL Wei, WH Lee, ZD Zhang, S Hartman, O Emanuelsson, V Stolc, S Weissman, MB Gerstein, Y Ruan, M Snyder (2007). "Mapping of transcription factor binding regions in mammalian cells by ChIP: comparison of array- and sequencing-based technologies." *Genome Res* 17: 898-909.
- S Washietl, JS Pedersen, JO Korbelt, C Stocsits, AR Gruber, J Hackermüller, J Hertel, M Lindemeyer, K Reiche, A Tanzer, C Ucla, C Wyss, SE Antonarakis, F Denoeud, J Lagarde, J Drenkow, P Kapranov, TR Gingeras, R Guigó, M Snyder, MB Gerstein, A Reymond, IL Hofacker, PF Stadler (2007). "Structured RNAs in the ENCODE selected regions of the human genome." *Genome Res* 17: 852-64.
- D Zheng, A Frankish, R Baertsch, P Kapranov, A Reymond, SW Choo, Y Lu, F Denoeud, SE Antonarakis, M Snyder, Y Ruan, CL Wei, TR Gingeras, R Guigó, J Harrow, MB Gerstein (2007). "Pseudogenes in the ENCODE regions: consensus annotation, analysis of transcription, and evolution." *Genome Res* 17: 839-51.
- ZD Zhang, A Paccanaro, Y Fu, S Weissman, Z Weng, J Chang, M Snyder, MB Gerstein (2007). "Statistical analysis of the genomic distribution and correlation of regulatory elements in the ENCODE regions." *Genome Res* 17: 787-97.
- JS Rozowsky, D Newburger, F Sayward, J Wu, G Jordan, JO Korbelt, U Nagalakshmi, J Yang, D Zheng, R Guigó, TR Gingeras, S Weissman, P Miller, M Snyder, MB Gerstein (2007). "The DART classification of unannotated transcription within the ENCODE regions: associating transcription with known and novel loci." *Genome Res* 17: 732-45.
- ND Trinklein, U Karaöz, J Wu, A Halees, S Force Aldred, PJ Collins, D Zheng, ZD Zhang, MB Gerstein, M Snyder, RM Myers, Z Weng (2007). "Integrated analysis of experimental data sets reveals many novel promoters in 1% of the human genome." *Genome Res* 17: 720-31.
- MB Gerstein, C Bruce, JS Rozowsky, D Zheng, J Du, JO Korbelt, O Emanuelsson, ZD Zhang, S Weissman, M Snyder (2007). "What is a gene, post-ENCODE? History and updated definition." *Genome Res* 17: 669-81.
- TE Royce, NJ Carriero, MB Gerstein (2007). "An efficient pseudomedian filter for tiling microarrays." *BMC Bioinformatics* 8: 186.
- JO Korbelt, AE Urban, F Grubert, J Du, TE Royce, P Starr, G Zhong, BS Emanuel, SM Weissman, M Snyder, MB Gerstein (2007). "Systematic prediction and validation of breakpoints associated with copy-number variants in the human genome." *Proc Natl Acad Sci U S A* 104: 10110-5.
- H Yu, R Jansen, G Stolovitzky, M Gerstein (2007). "Total ancestry measure: quantifying the similarity in tree-like classification, with genomic applications." *Bioinformatics* 23: 2163-73.
- L Wu, SI Hwang, K Rezaul, LJ Lu, V Mayya, M Gerstein, JK Eng, DH Lundgren, DK Han (2007). "Global survey of human T leukemic cells by integrating proteomics and transcriptomics profiling." *Mol Cell Proteomics* 6: 1343-53.
- SC Flores, LJ Lu, J Yang, N Carriero, MB Gerstein (2007). "Hinge Atlas: relating protein sequence to sites of structural flexibility." *BMC Bioinformatics* 8: 167.
- ZD Zhang, J Rozowsky, HY Lam, J Du, M Snyder, M Gerstein (2007). "Telescope: online analysis pipeline for high-density tiling microarray data." *Genome Biol* 8: R81.

- M Gerstein, M Seringhaus, S Fields (2007). "Structured digital abstract makes text mining easy." *Nature* 447: 142.
- AK Smith, KH Cheung, KY Yip, M Schultz, MK Gerstein (2007). "LinkHub: a Semantic Web system that facilitates cross-database queries and information retrieval in proteomics." *BMC Bioinformatics* 8 Suppl 3: S5.
- X Zhu, M Gerstein, M Snyder (2007). "Getting connected: analysis and principles of biological networks." *Genes Dev* 21: 1010-24.
- M Gerstein, SM Douglas (2007). "RNAi development." *PLoS Comput Biol* 3: e80.
- H Yu, PM Kim, E Sprecher, V Trifonov, M Gerstein (2007). "The importance of bottlenecks in protein networks: correlation with gene essentiality and expression dynamics." *PLoS Comput Biol* 3: e59.
- TE Royce, JS Rozowsky, MB Gerstein (2007). "Assessing the need for sequence-based normalization in tiling microarray experiments." *Bioinformatics* 23: 988-97.
- D Zheng, MB Gerstein (2007). "The ambiguous boundary between genes and pseudogenes: the dead rise up, or do they?" *Trends Genet* 23: 219-24.
- L Li, X Wang, R Sasidharan, V Stolc, W Deng, H He, J Korbel, X Chen, W Tongprasit, P Ronald, R Chen, M Gerstein, XW Deng (2007). "Global identification and characterization of transcriptionally active regions in the rice genome." *PLoS One* 2: e294.
- SC Popescu, GV Popescu, S Bachan, Z Zhang, M Seay, M Gerstein, M Snyder, SP Dinesh-Kumar (2007). "Differential binding of calmodulin-related proteins to their targets revealed through high-density Arabidopsis protein microarrays." *Proc Natl Acad Sci U S A* 104: 4730-5.
- MG Smith, TA Gianoulis, S Pukatzki, JJ Mekalanos, LN Ornston, M Gerstein, M Snyder (2007). "New insights into *Acinetobacter baumannii* pathogenesis revealed by high-density pyrosequencing and transposon mutagenesis." *Genes Dev* 21: 601-14.
- Z Zhang, AW Pang, M Gerstein (2007). "Comparative analysis of genome tiling array data reveals many novel primate-specific functional RNAs in human." *BMC Evol Biol* 7 Suppl 1: S14.
- MR Seringhaus, MB Gerstein (2007). "Publishing perishing? Towards tomorrow's information architecture." *BMC Bioinformatics* 8: 17.
- M Seringhaus, M Gerstein (2007). "Chemistry Nobel rich in structure." *Science* 315: 40-1.
- H Yu, K Nguyen, T Royce, J Qian, K Nelson, M Snyder, M Gerstein (2007). "Positional artifacts in microarrays: experimental verification and construction of COP, an automated detection tool." *Nucleic Acids Res* 35: e8.
- O Emanuelsson, U Nagalakshmi, D Zheng, JS Rozowsky, AE Urban, J Du, Z Lian, V Stolc, S Weissman, M Snyder, MB Gerstein (2007). "Assessing the performance of different high-density tiling microarray strategies for mapping transcribed regions of the human genome." *Genome Res* 17: 886-97.
- JE Karro, Y Yan, D Zheng, Z Zhang, N Carriero, P Cayting, P Harrison, M Gerstein (2007). "Pseudogene.org: a comprehensive database and comparison platform for pseudogene annotation." *Nucleic Acids Res* 35: D55-60.
- M Gerstein, D Greenbaum, K Cheung, PL Miller (2007). "An interdepartmental Ph.D. program in computational biology and bioinformatics: the Yale perspective." *J Biomed Inform* 40: 73-9.

-- 2006 --

The Death of the Scientific Paper. Seringhaus M, Gerstein M (2006). *The Scientist*. 20(9): 25. Analytical Evolutionary Model for Protein Fold Occurrence in Genomes, Accounting for the Effects of Gene Duplication, Deletion, Acquisition and Selective Pressure. M Kamal, N Luscombe, J Qian, M Gerstein (2006) in *Power Laws, Scale-Free Networks and Genome Biology* (edited by EV Koonin, YI Wolf, GP Karev; Springer, New York), pages 165-193.

- J Rozowsky, J Wu, Z Lian, U Nagalakshmi, JO Korbelt, P Kapranov, D Zheng, S Dyke, P Newburger, P Miller, TR Gingeras, S Weissman, M Gerstein, M Snyder (2006). "Novel transcribed regions in the human genome." *Cold Spring Harb Symp Quant Biol* 71: 111-6.
- PM Kim, LJ Lu, Y Xia, MB Gerstein (2006). "Relating three-dimensional structures to protein networks provides evolutionary insights." *Science* 314: 1938-41.
- A Smith, M Gerstein (2006). "Data mining on the web." *Science* 314: 1682; author reply 1682.
- Y Liu, J Li, L Sam, CS Goh, M Gerstein, YA Lussier (2006). "An integrative genomic approach to uncover molecular mechanisms of prokaryotic traits." *PLoS Comput Biol* 2: e159.
- X Zhu, M Gerstein, M Snyder (2006). "ProCAT: a data analysis approach for protein microarrays." *Genome Biol* 7: R110.
- LY Wang, M Snyder, M Gerstein (2006). "BoCaTFBS: a boosted cascade learner to refine the binding sites suggested by CHIP-chip experiments." *Genome Biol* 7: R102.
- AE Burba, U Lehnert, EZ Yu, M Gerstein (2006). "Helix Interaction Tool (HIT): a web-based tool for analysis of helix-helix interactions in proteins." *Bioinformatics* 22: 2735-8.
- J Du, JS Rozowsky, JO Korbelt, ZD Zhang, TE Royce, MH Schultz, M Snyder, M Gerstein (2006). "A supervised hidden markov model framework for efficiently segmenting tiling array data in transcriptional and chIP-chip experiments: systematically incorporating validated biological knowledge." *Bioinformatics* 22: 3016-24.
- CS Goh, TA Gianoulis, Y Liu, J Li, A Paccanaro, YA Lussier, M Gerstein (2006). "Integration of curated databases to identify genotype-phenotype associations." *BMC Genomics* 7: 257.
- KY Yip, H Yu, PM Kim, M Schultz, M Gerstein (2006). "The tYNA platform for comparative interactomics: a web tool for managing, comparing and mining multiple networks." *Bioinformatics* 22: 2968-70.
- H Yu, M Gerstein (2006). "Genomic analysis of the hierarchical structure of regulatory networks." *Proc Natl Acad Sci U S A* 103: 14724-31.
- T Srikantha, AR Borneman, KJ Daniels, C Pujol, W Wu, MR Seringhaus, M Gerstein, S Yi, M Snyder, DR Soll (2006). "TOS9 regulates white-opaque switching in *Candida albicans*." *Eukaryot Cell* 5: 1674-87.
- TE Royce, JS Rozowsky, NM Luscombe, O Emanuelsson, H Yu, X Zhu, M Snyder, MB Gerstein (2006). "Extrapolating traditional DNA microarray statistics to tiling and protein microarray technologies." *Methods Enzymol* 411: 282-311.
- R Pinard, A de Winter, GJ Sarkis, MB Gerstein, KR Tartaro, RN Plant, M Egholm, JM Rothberg, JH Leamon (2006). "Assessment of whole genome amplification-induced bias through high-throughput, massively parallel whole genome sequencing." *BMC Genomics* 7: 216.
- D Zheng, MB Gerstein (2006). "A computational approach for identifying pseudogenes in the ENCODE regions." *Genome Biol* 7 Suppl 1: S13.1-10.
- M Seringhaus, A Paccanaro, A Borneman, M Snyder, M Gerstein (2006). "Predicting essential genes in fungal genomes." *Genome Res* 16: 1126-35.
- M Gerstein, D Zheng (2006). "The real life of pseudogenes." *Sci Am* 295: 48-55.
- H Yu, Y Xia, V Trifonov, M Gerstein (2006). "Design principles of molecular networks revealed by global comparisons and composite motifs." *Genome Biol* 7: R55.
- NR Voss, M Gerstein, TA Steitz, PB Moore (2006). "The geometry of the ribosomal polypeptide exit tunnel." *J Mol Biol* 360: 893-906.
- M Seringhaus, A Kumar, J Hartigan, M Snyder, M Gerstein (2006). "Genomic analysis of insertion behavior and target specificity of mini-Tn7 and Tn3 transposons in *Saccharomyces cerevisiae*." *Nucleic Acids Res* 34: e57.

- M Gerstein (2006). "Tools needed to navigate landscape of the genome." *Nature* 440: 740.
- Z Zhang, N Carriero, D Zheng, J Karro, PM Harrison, M Gerstein (2006). "PseudoPipe: an automated pseudogene identification pipeline." *Bioinformatics* 22: 1437-9.
- NJ Krogan, G Cagney, H Yu, G Zhong, X Guo, A Ignatchenko, J Li, S Pu, N Datta, AP Tikuisis, T Punna, JM Peregrín-Alvarez, M Shales, X Zhang, M Davey, MD Robinson, A Paccanaro, JE Bray, A Sheung, B Beattie, DP Richards, V Canadien, A Lalev, F Mena, P Wong, A Starostine, MM Canete, J Vlasblom, S Wu, C Orsi, SR Collins, S Chandran, R Haw, JJ Rilstone, K Gandi, NJ Thompson, G Musso, P St Onge, S Ghanny, MH Lam, G Butland, AM Altaf-Ul, S Kanaya, A Shilatifard, E O'Shea, JS Weissman, CJ Ingles, TR Hughes, J Parkinson, M Gerstein, SJ Wodak, A Emili, JF Greenblatt (2006). "Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*." *Nature* 440: 637-43.
- AE Urban, JO Korbelt, R Selzer, T Richmond, A Hacker, GV Popescu, JF Cubells, R Green, BS Emanuel, MB Gerstein, SM Weissman, M Snyder (2006). "High-resolution mapping of DNA copy alterations in human chromosome 22 using high-density tiling oligonucleotide arrays." *Proc Natl Acad Sci U S A* 103: 4534-9.
- H Yu, A Paccanaro, V Trifonov, M Gerstein (2006). "Predicting interactions in protein networks by completing defective cliques." *Bioinformatics* 22: 823-9.
- AR Borneman, JA Leigh-Bell, H Yu, P Bertone, M Gerstein, M Snyder (2006). "Target hub proteins serve as master regulators of development in yeast." *Genes Dev* 20: 435-48.
- Y Xia, LJ Lu, M Gerstein (2006). "Integrated prediction of the helical membrane protein interactome in yeast." *J Mol Biol* 357: 339-49.
- S Flores, N Echols, D Milburn, B Hespeneide, K Keating, J Lu, S Wells, EZ Yu, M Thorpe, M Gerstein (2006). "The Database of Macromolecular Motions: new features added at the decade mark." *Nucleic Acids Res* 34: D296-301.
- P Bertone, V Trifonov, JS Rozowsky, F Schubert, O Emanuelsson, J Karro, MY Kao, M Snyder, M Gerstein (2006). "Design optimization methods for genomic DNA tiling arrays." *Genome Res* 16: 271-81.

-- 2005 --

- Inferring Protein-Protein Interactions Using Interaction Network Topologies. A Paccanaro, V Trifonov, H Yu, M Gerstein (2005). International Joint Conference on Neural Networks (IJCNN, Jul. 31-Aug. 4, Montreal, Canada), pages 161 - 166, vol. 1. Protein Interaction Prediction by Integrating Genomic Features and Protein Interaction Network Analysis. LJ Lu, Y Xia, H Yu, A Rives, H Lu, F Schubert, M Gerstein (2005). *Data Analysis and Visualization in Genomics and Proteomics* (Wiley, NY).
- DM Gelperin, MA White, ML Wilkinson, Y Kon, LA Kung, KJ Wise, N Lopez-Hoyo, L Jiang, S Piccirillo, H Yu, M Gerstein, ME Dumont, EM Phizicky, M Snyder, EJ Grayhack (2005). "Biochemical and genetic analysis of the yeast proteome with a movable ORF collection." *Genes Dev* 19: 2816-26.
- J Ptacek, G Devgan, G Michaud, H Zhu, X Zhu, J Fasolo, H Guo, G Jona, A Breitkreutz, R Sopko, RR McCartney, MC Schmidt, N Rachidi, SJ Lee, AS Mah, L Meng, MJ Stark, DF Stern, C De Virgilio, M Tyers, B Andrews, M Gerstein, B Schweitzer, PF Predki, M Snyder (2005). "Global analysis of protein phosphorylation in yeast." *Nature* 438: 679-84.
- SE Hartman, P Bertone, AK Nath, TE Royce, M Gerstein, S Weissman, M Snyder (2005). "Global changes in STAT target selection and transcription regulation upon interferon treatments." *Genes Dev* 19: 2953-68.
- V Stolc, L Li, X Wang, X Li, N Su, W Tongprasit, B Han, Y Xue, J Li, M Snyder, M Gerstein, J Wang, XW Deng (2005). "A pilot study of transcription unit analysis in rice using oligonucleotide tiling-path microarray." *Plant Mol Biol* 59: 137-49.

- A Smith, D Greenbaum, SM Douglas, M Long, M Gerstein (2005). "Network security and data integrity in academia: an assessment and a proposal for large-scale archiving." *Genome Biol* 6: 119.
- SM Douglas, GT Montelione, M Gerstein (2005). "PubNet: a flexible system for visualizing literature derived networks." *Genome Biol* 6: R80.
- T Coric, D Zheng, M Gerstein, CM Canessa (2005). "Proton sensitivity of ASIC1 appeared with the rise of fishes by changes of residues in the region that follows TM1 in the ectodomain of the channel." *J Physiol* 568: 725-35.
- LJ Lu, Y Xia, A Paccanaro, H Yu, M Gerstein (2005). "Assessing the limits of genomic data integration for predicting protein networks." *Genome Res* 15: 945-53.
- TE Royce, JS Rozowsky, P Bertone, M Samanta, V Stolc, S Weissman, M Snyder, M Gerstein (2005). "Issues in the analysis of oligonucleotide tiling microarrays for transcript mapping." *Trends Genet* 21: 466-75.
- KH Cheung, KY Yip, A Smith, R Deknikker, A Masiar, M Gerstein (2005). "YeastHub: a semantic web use case for integrating data in the life sciences domain." *Bioinformatics* 21 Suppl 1: i85-96.
- D Zheng, Z Zhang, PM Harrison, J Karro, N Carriero, M Gerstein (2005). "Integrated pseudogene annotation for human chromosome 22: evidence for transcription." *J Mol Biol* 349: 27-45.
- P Bertone, M Gerstein, M Snyder (2005). "Applications of DNA tiling arrays to experimental genome annotation and regulatory pathway discovery." *Chromosome Res* 13: 259-74.
- Y Gilad, SA Rifkin, P Bertone, M Gerstein, KP White (2005). "Multi-species microarrays reveal the effect of sequence divergence on gene expression profiles." *Genome Res* 15: 674-80.
- PM Harrison, D Zheng, Z Zhang, N Carriero, M Gerstein (2005). "Transcribed processed pseudogenes in the human genome: an intermediate form of expressed retrosequence lacking protein-coding ability." *Nucleic Acids Res* 33: 2374-83.
- D Huber, D Boyd, Y Xia, MH Olma, M Gerstein, J Beckwith (2005). "Use of thioredoxin as a reporter to identify a subset of Escherichia coli signal sequences that promote signal recognition particle-dependent translocation." *J Bacteriol* 187: 2983-91.
- TB Acton, KC Gunsalus, R Xiao, LC Ma, J Aramini, MC Baran, YW Chiang, T Climent, B Cooper, NG Denissova, SM Douglas, JK Everett, CK Ho, D Macapagal, PK Rajan, R Shastry, LY Shih, GV Swapna, M Wilson, M Wu, M Gerstein, M Inouye, JF Hunt, GT Montelione (2005). "Robotic cloning and Protein Production Platform of the Northeast Structural Genomics Consortium." *Methods Enzymol* 394: 210-43.
- S Balasubramanian, Y Xia, E Freinkman, M Gerstein (2005). "Sequence variation in G-protein-coupled receptors: analysis of single nucleotide polymorphisms." *Nucleic Acids Res* 33: 1710-21.
- H Grosshans, T Johnson, KL Reinert, M Gerstein, FJ Slack (2005). "The temporal patterning microRNA let-7 regulates several transcription factors at the larval to adult transition in *C. elegans*." *Dev Cell* 8: 321-30.
- V Alexandrov, U Lehnert, N Echols, D Milburn, D Engelman, M Gerstein (2005). "Normal modes for predicting protein motions: a comprehensive database assessment and associated Web tool." *Protein Sci* 14: 633-43.
- NR Voss, M Gerstein (2005). "Calculation of standard atomic volumes for RNA and comparison with proteins: RNA is packed more tightly." *J Mol Biol* 346: 477-92.
- D Greenbaum, A Smith, M Gerstein (2005). "Impediments to database interoperation: legal issues and security concerns." *Nucleic Acids Res* 33: D3-4.
- N Carriero, MV Osier, KH Cheung, PL Miller, M Gerstein, H Zhao, B Wu, S Rifkin, J Chang, H Zhang, K White, K Williams, M Schultz (2005). "A high productivity/low maintenance approach to high-performance computation for biomedicine: four case studies." *J Am Med Inform Assoc* 12: 90-8.

-- 2004 --

- An XML-Based Approach to Integrating Heterogeneous Yeast Genome Data. KH Cheung, D Pan, A Smith, M Seringhaus, SM Douglas, M Gerstein. 2004 International Conference on Mathematics and Engineering Techniques in Medicine and Biological Sciences (METMBS); pp 236-242.
- U Lehnert, Y Xia, TE Royce, CS Goh, Y Liu, A Senes, H Yu, ZL Zhang, DM Engelman, M Gerstein (2004). "Computational analysis of membrane proteins: genomic occurrence, structure prediction and helix interactions." *Q Rev Biophys* 37: 121-46.
- EJ White, O Emanuelsson, D Scalzo, T Royce, S Kosak, EJ Oakeley, S Weissman, M Gerstein, M Groudine, M Snyder, D Schübeler (2004). "DNA replication-timing analysis of human chromosome 22 at high resolution and different developmental states." *Proc Natl Acad Sci U S A* 101: 17771-6.
- P Berman, P Bertone, B Dasgupta, M Gerstein, MY Kao, M Snyder (2004). "Fast optimal genome tiling with applications to microarray design and homology search." *J Comput Biol* 11: 766-85.
- P Bertone, V Stolc, TE Royce, JS Rozowsky, AE Urban, X Zhu, JL Rinn, W Tongprasit, M Samanta, S Weissman, M Gerstein*, M Snyder* (2004). "Global identification of human transcribed sequences with genome tiling arrays." *Science* 306: 2242-6.
- ENCODE Project Consortium (2004). "The ENCODE (ENCyclopedia Of DNA Elements) Project." *Science* 306: 636-40.
- N Lin, B Wu, R Jansen, M Gerstein, H Zhao (2004). "Information assessment on predicting protein-protein interactions." *BMC Bioinformatics* 5: 154.
- DA Hall, H Zhu, X Zhu, T Royce, M Gerstein, M Snyder (2004). "Regulation of gene expression by a metabolic enzyme." *Science* 306: 482-4.
- A Kumar, M Seringhaus, MC Biery, RJ Sarnovsky, L Umansky, S Piccirillo, M Heidtman, KH Cheung, CJ Dobry, MB Gerstein, NL Craig, M Snyder (2004). "Large-scale mutagenesis of the yeast genome using a Tn7-derived multipurpose transposon." *Genome Res* 14: 1975-86.
- R Jansen, M Gerstein (2004). "Analyzing protein function on a genomic scale: the importance of gold-standard positives and negatives for network prediction." *Curr Opin Microbiol* 7: 535-45.
- NM Luscombe, MM Babu, H Yu, M Snyder, SA Teichmann, M Gerstein (2004). "Genomic analysis of regulatory network dynamics reveals large topological changes." *Nature* 431: 308-12.
- Y Liu, PM Harrison, V Kunin, M Gerstein (2004). "Comprehensive analysis of pseudogenes in prokaryotes: widespread gene decay and failure of putative horizontally transferred genes." *Genome Biol* 5: R64.
- Z Zhang, M Gerstein (2004). "Large-scale analysis of pseudogenes in the human genome." *Curr Opin Genet Dev* 14: 328-35.
- Z Wunderlich, TB Acton, J Liu, G Kornhaber, J Everett, P Carter, N Lan, N Echols, M Gerstein, B Rost, GT Montelione (2004). "The protein target list of the Northeast Structural Genomics Consortium." *Proteins* 56: 181-7.
- MM Babu, NM Luscombe, L Aravind, M Gerstein, SA Teichmann (2004). "Structure and evolution of transcriptional regulatory networks." *Curr Opin Struct Biol* 14: 283-91.
- Y Xia, H Yu, R Jansen, M Seringhaus, S Baxter, D Greenbaum, H Zhao, M Gerstein (2004). "Analyzing cellular biochemistry in terms of molecular networks." *Annu Rev Biochem* 73: 1051-87.
- JL Rinn, JS Rozowsky, IJ Laurenzi, PH Petersen, K Zou, W Zhong, M Gerstein, M Snyder (2004). "Major molecular differences between mammalian sexes are involved in drug metabolism and renal function." *Dev Cell* 6: 791-800.
- D Greenbaum, SM Douglas, A Smith, J Lim, M Fischer, M Schultz, M Gerstein (2004). "Computer security in academia-a potential roadblock to distributed annotation of the human genome." *Nat Biotechnol* 22: 771-2.

- H Yu, NM Luscombe, HX Lu, X Zhu, Y Xia, JD Han, N Bertin, S Chung, M Vidal, M Gerstein (2004). "Annotation transfer between genomes: protein-protein interologs and protein-DNA regulogs." *Genome Res* 14: 1107-18.
- H Yu, D Greenbaum, H Xin Lu, X Zhu, M Gerstein (2004). "Genomic analysis of essentiality within protein networks." *Trends Genet* 20: 227-31.
- LL Freeman-Cook, AM Dixon, JB Frank, Y Xia, L Ely, M Gerstein, DM Engelman, D DiMaio (2004). "Selection and characterization of small random transmembrane proteins that bind and activate the platelet-derived growth factor beta receptor." *J Mol Biol* 338: 907-20.
- CS Goh, D Milburn, M Gerstein (2004). "Conformational changes associated with protein-protein interactions." *Curr Opin Struct Biol* 14: 104-9.
- G Euskirchen, TE Royce, P Bertone, R Martone, JL Rinn, FK Nelson, F Sayward, NM Luscombe, P Miller, M Gerstein, S Weissman, M Snyder (2004). "CREB binds to multiple loci on human chromosome 22." *Mol Cell Biol* 24: 3804-14.
- R Das, M Gerstein (2004). "A method using active-site sequence conservation to find functional shifts in protein families: application to the enzymes of central metabolism, leading to the identification of an anomalous isocitrate dehydrogenase in pathogens." *Proteins* 55: 455-63.
- M Gerstein, N Echols (2004). "Exploring the range of protein flexibility, from a structural proteomics perspective." *Curr Opin Chem Biol* 8: 14-9.
- Y Liu, M Gerstein, DM Engelman (2004). "Transmembrane protein domains rarely use covalent domain recombination as an evolutionary mechanism." *Proc Natl Acad Sci U S A* 101: 3495-7.
- Z Zhang, N Carriero, M Gerstein (2004). "Comparative analysis of processed pseudogenes in the mouse and human genomes." *Trends Genet* 20: 62-7.
- CS Goh, N Lan, SM Douglas, B Wu, N Echols, A Smith, D Milburn, GT Montelione, H Zhao, M Gerstein (2004). "Mining the structural genomics pipeline: identification of protein properties that affect high-throughput experimental analysis." *J Mol Biol* 336: 115-30.
- H Yu, X Zhu, D Greenbaum, J Karro, M Gerstein (2004). "TopNet: a tool for comparing biological sub-networks, correlating protein properties with topological statistics." *Nucleic Acids Res* 32: 328-37.
- V Alexandrov, M Gerstein (2004). "Using 3D Hidden Markov Models that explicitly represent spatial coordinates to model and compare protein structures." *BMC Bioinformatics* 5: 2.
- S Li, CM Armstrong, N Bertin, H Ge, S Milstein, M Boxem, PO Vidalain, JD Han, A Chesneau, T Hao, DS Goldberg, N Li, M Martinez, JF Rual, P Lamesch, L Xu, M Tewari, SL Wong, LV Zhang, GF Berriz, L Jacotot, P Vaglio, J Reboul, T Hirozane-Kishikawa, Q Li, HW Gabel, A Elewa, B Baumgartner, DJ Rose, H Yu, S Bosak, R Sequerra, A Fraser, SE Mango, WM Saxton, S Strome, S Van Den Heuvel, F Piano, J Vandenhoute, C Sardet, M Gerstein, L Doucette-Stamm, KC Gunsalus, JW Harper, ME Cusick, FP Roth, DE Hill, M Vidal (2004). "A map of the interactome network of the metazoan *C. elegans*." *Science* 303: 540-3.

-- 2003 --

- WG Krebs, J Tsai, V Alexandrov, J Junker, R Jansen, M Gerstein (2003). "Tools and databases to analyze protein flexibility; approaches to mapping implied features onto sequences." *Methods Enzymol* 374: 544-84.
- D Greenbaum, J Lim, M Gerstein (2003). "An analysis of the present system of scientific publishing: what's wrong and where to go from here" *Interdiscip Sci Rev* 28: 293-302.
- Z Lian, G Euskirchen, J Rinn, R Martone, P Bertone, S Hartman, T Royce, K Nelson, F Sayward, N Luscombe, J Yang, JL Li, P Miller, AE Urban, M Gerstein, S Weissman, M Snyder (2003).

- "Identification of novel functional elements in the human genome." *Cold Spring Harb Symp Quant Biol* 68: 317-22.
- Y Kluger, H Yu, J Qian, M Gerstein (2003). "Relationship between gene co-expression and probe localization on microarray slides." *BMC Genomics* 4: 49.
- Z Zhang, PM Harrison, Y Liu, M Gerstein (2003). "Millions of years of evolution preserved: a comprehensive catalog of the processed pseudogenes in the human genome." *Genome Res* 13: 2541-58.
- Y Jiao, H Yang, L Ma, N Sun, H Yu, T Liu, Y Gao, H Gu, Z Chen, M Wada, M Gerstein, H Zhao, LJ Qu, XW Deng (2003). "A genome-wide analysis of blue-light regulation of Arabidopsis transcription factor gene expression during seedling development." *Plant Physiol* 133: 1480-93.
- Z Zhang, M Gerstein (2003). "Reconstructing genetic networks in yeast." *Nat Biotechnol* 21: 1295-7.
- PM Harrison, N Carrero, Y Liu, M Gerstein (2003). "A "polyORFomic" analysis of prokaryote genomes using disabled-homology filtering reveals conserved but undiscovered short ORFs." *J Mol Biol* 333: 885-92.
- R Jansen, H Yu, D Greenbaum, Y Kluger, NJ Krogan, S Chung, A Emili, M Snyder, JF Greenblatt, M Gerstein (2003). "A Bayesian networks approach for predicting protein-protein interactions from genomic data." *Science* 302: 449-53.
- J Qian, J Lin, NM Luscombe, H Yu, M Gerstein (2003). "Prediction of regulatory networks: genome-wide identification of transcription factor targets from gene expression data." *Bioinformatics* 19: 1917-26.
- R Martone, G Euskirchen, P Bertone, S Hartman, TE Royce, NM Luscombe, JL Rinn, FK Nelson, P Miller, M Gerstein, S Weissman, M Snyder (2003). "Distribution of NF-kappaB-binding sites across human chromosome 22." *Proc Natl Acad Sci U S A* 100: 12247-52.
- Z Zhang, M Gerstein (2003). "Patterns of nucleotide substitution, insertion and deletion in the human genome inferred from pseudogenes." *Nucleic Acids Res* 31: 5338-48.
- D Greenbaum, C Colangelo, K Williams, M Gerstein (2003). "Comparing protein abundance and mRNA expression levels on a genomic scale." *Genome Biol* 4: 117.
- D Greenbaum, M Gerstein (2003). "A universal legal framework as a prerequisite for database interoperability." *Nat Biotechnol* 21: 979-82.
- Z Zhang, M Gerstein (2003). "The human genome has 49 cytochrome c pseudogenes, including a relic of a primordial gene that still functions in mouse." *Gene* 312: 61-72.
- H Yu, NM Luscombe, J Qian, M Gerstein (2003). "Genomic analysis of gene expression relationships in transcriptional regulatory networks." *Trends Genet* 19: 422-7.
- J Qian, Y Kluger, H Yu, M Gerstein (2003). "Identification and correction of spurious spatial correlations in microarray data." *Biotechniques* 35: 42-4, 46, 48.
- M Gerstein, JM Thornton (2003). "Sequences and topology." *Curr Opin Struct Biol* 13: 341-3.
- NM Luscombe, TE Royce, P Bertone, N Echols, CE Horak, JT Chang, M Snyder, M Gerstein (2003). "ExpressYourself: A modular platform for processing and visualizing microarray data." *Nucleic Acids Res* 31: 3477-82.
- Z Zhang, M Gerstein (2003). "Of mice and men: phylogenetic footprinting aids the discovery of regulatory elements." *J Biol* 2: 11.
- PM Harrison, M Gerstein (2003). "A method to assess compositional bias in biological sequences and its application to prion-like glutamine/asparagine-rich domains in eukaryotic proteomes." *Genome Biol* 4: R40.
- MS Kimber, F Vallee, S Houston, A Necakov, T Skarina, E Evdokimova, S Beasley, D Christendat, A Savchenko, CH Arrowsmith, M Vedadi, M Gerstein, AM Edwards (2003). "Data mining

crystallization databases: knowledge-based approaches to optimize protein crystal screens." *Proteins* 51: 562-8.

CS Goh, N Lan, N Echols, SM Douglas, D Milburn, P Bertone, R Xiao, LC Ma, D Zheng, Z Wunderlich, T Acton, GT Montelione, M Gerstein (2003). "SPINE 2: a system for collaborative structural proteomics within a federated database framework." *Nucleic Acids Res* 31: 2833-8.

Z Zhang, M Gerstein (2003). "Identification and characterization of over 100 mitochondrial ribosomal protein pseudogenes in the human genome." *Genomics* 81: 468-80.

M Snyder, M Gerstein (2003). "Genomics. Defining genes in the genomics era." *Science* 300: 258-60.

R Jansen, HJ Bussemaker, M Gerstein (2003). "Revisiting the codon adaptation index from a whole-genome perspective: analyzing the relationship between gene expression and codon occurrence in yeast using a variety of models." *Nucleic Acids Res* 31: 2242-51.

Y Kluger, R Basri, JT Chang, M Gerstein (2003). "Spectral biclustering of microarray data: coclustering genes and conditions." *Genome Res* 13: 703-16.

M Gerstein, A Edwards, CH Arrowsmith, GT Montelione (2003). "Structural genomics: current progress." *Science* 299: 1663.

JL Rinn, G Euskirchen, P Bertone, R Martone, NM Luscombe, S Hartman, PM Harrison, FK Nelson, P Miller, M Gerstein, S Weissman, M Snyder (2003). "The transcriptional activity of human Chromosome 22." *Genes Dev* 17: 529-40.

PM Harrison, D Milburn, Z Zhang, P Bertone, M Gerstein (2003). "Identification of pseudogenes in the *Drosophila melanogaster* genome." *Nucleic Acids Res* 31: 1033-7.

A Savchenko, A Yee, A Khachatryan, T Skarina, E Evdokimova, M Pavlova, A Semesi, J Northey, S Beasley, N Lan, R Das, M Gerstein, CH Arrowsmith, AM Edwards (2003). "Strategies for structural proteomics of prokaryotes: Quantifying the advantages of studying orthologous proteins and of using both NMR and X-ray crystallography approaches." *Proteins* 50: 392-9.

N Lan, GT Montelione, M Gerstein (2003). "Ontologies for proteomics: towards a systematic definition of structure and function that scales to the genome level." *Curr Opin Chem Biol* 7: 44-54.

N Echols, D Milburn, M Gerstein (2003). "MolMovDB: analysis and visualization of conformational change and structural flexibility." *Nucleic Acids Res* 31: 478-82.

-- 2002 --

J Tsai, M Gerstein (2002). "Calculations of protein volumes: sensitivity analysis and parameter database." *Bioinformatics* 18: 985-95.

A Kumar, S Agarwal, JA Heyman, S Matson, M Heidtman, S Piccirillo, L Umansky, A Drawid, R Jansen, Y Liu, KH Cheung, P Miller, M Gerstein, GS Roeder, M Snyder (2002). "Subcellular localization of the yeast proteome." *Genes Dev* 16: 707-19.

D Greenbaum, R Jansen, M Gerstein (2002). "Analysis of mRNA expression and protein abundance data: an approach for the comparison of the enrichment of features in the cellular population of proteins and transcripts." *Bioinformatics* 18: 585-96.

H Hegyi, J Lin, D Greenbaum, M Gerstein (2002). "Structural genomics analysis: characteristics of atypical, common, and horizontally transferred folds." *Proteins* 47: 126-41.

P Harrison, A Kumar, N Lan, N Echols, M Snyder, M Gerstein (2002). "A small reservoir of disabled ORFs in the yeast genome and its implications for the dynamics of proteome evolution." *J Mol Biol* 316: 409-19.

PM Harrison, M Gerstein (2002). "Studying genomes through the aeons: protein families, pseudogenes and proteome evolution." *J Mol Biol* 318: 1155-74.

- PM Harrison, A Kumar, N Lang, M Snyder, M Gerstein (2002). "A question of size: the eukaryotic proteome and the problems in defining it." *Nucleic Acids Res* 30: 1083-90.
- PM Harrison, H Hegyi, S Balasubramanian, NM Luscombe, P Bertone, N Echols, T Johnson, M Gerstein (2002). "Molecular fossils in the human genome: identification and analysis of the pseudogenes in chromosomes 21 and 22." *Genome Res* 12: 272-80.
- R Jansen, D Greenbaum, M Gerstein (2002). "Relating whole-genome expression data with protein-protein interactions." *Genome Res* 12: 37-46.
- M Gerstein, N Lan, R Jansen (2002). "Proteomics. Integrating interactomes." *Science* 295: 284-7.
- Towards a systematic definition of protein function that scales to the genome level: Defining function in terms of interactions. N Lan, R Jansen, M Gerstein (2002). Proceedings of the IEEE 90:1848-1858 .
- M Gerstein, J Junker (2002). "Blurring the boundaries between scientific 'papers' and biological databases" *Nature Yearbook of Science and Technology* : 210-212 (ed. D Butler, Palgrave Macmillan Publishers).
- Fast optimal genome tiling with applications to microarray design and homology search. P Berman, P Bertone, B DasGupta, M Gerstein, M-Y Kao, M Snyder. (2002) Proceedings of the 2nd International Workshop on Algorithms in Bioinformatics. Springer-Verlag LNCS 2452: 419-433 .
- R Jansen, N Lan, J Qian, M Gerstein (2002). "Integration of genomic datasets to predict protein complexes in yeast." *J Struct Funct Genomics* 2: 71-81.
- CE Horak, NM Luscombe, J Qian, P Bertone, S Piccirillo, M Gerstein, M Snyder (2002). "Complex transcriptional circuitry at the G1/S transition in *Saccharomyces cerevisiae*." *Genes Dev* 16: 3017-33.
- KH Cheung, K White, J Hager, M Gerstein, V Reinke, K Nelson, P Masiar, R Srivastava, Y Li, J Li, H Zhao, J Li, DB Allison, M Snyder, P Miller, K Williams (2002). "YMD: a microarray database for large-scale gene expression analysis." *Proc AMIA Symp* : 140-4.
- D Schneider, Y Liu, M Gerstein, DM Engelman (2002). "Thermostability of membrane protein helix-helix interaction elucidated by statistical analysis." *FEBS Lett* 532: 231-6.
- A Mateos, J Dopazo, R Jansen, Y Tu, M Gerstein, G Stolovitzky (2002). "Systematic learning of gene functional classes from DNA array expression data by using multilayer perceptrons." *Genome Res* 12: 1703-15.
- ZL Zhang, PM Harrison, M Gerstein (2002). "Digging deep for ancient relics: a survey of protein motifs in the intergenic sequences of four eukaryotic genomes." *J Mol Biol* 323: 811-22.
- J Lin, J Qian, D Greenbaum, P Bertone, R Das, N Echols, A Senes, B Stenger, M Gerstein (2002). "GeneCensus: genome comparisons in terms of metabolic pathway activity and protein family sharing." *Nucleic Acids Res* 30: 4574-82.
- Z Lian, Y Kluger, DS Greenbaum, D Tuck, M Gerstein, N Berliner, SM Weissman, PE Newburger (2002). "Genomic and proteomic analysis of the myeloid differentiation program: global analysis of gene expression during induced differentiation in the MPRO cell line." *Blood* 100: 3209-20.
- Y Liu, DM Engelman, M Gerstein (2002). "Genomic analysis of membrane protein families: abundance and conserved motifs." *Genome Biol* 3: research0054.
- Z Zhang, P Harrison, M Gerstein (2002). "Identification and analysis of over 2000 ribosomal protein pseudogenes in the human genome." *Genome Res* 12: 1466-82.
- AM Edwards, B Kus, R Jansen, D Greenbaum, J Greenblatt, M Gerstein (2002). "Bridging structural biology and genomics: assessing protein interaction data with known complexes." *Trends Genet* 18: 529-36.
- WG Krebs, V Alexandrov, CA Wilson, N Echols, H Yu, M Gerstein (2002). "Normal mode analysis of macromolecular motions in a database framework: developing mode concentration as a useful classifying statistic." *Proteins* 48: 682-95.

- NM Luscombe, J Qian, Z Zhang, T Johnson, M Gerstein (2002). "The dominance of the population by a selected few: power-law behaviour applies to a wide variety of genomic properties." *Genome Biol* 3: RESEARCH0040.
- G Giaever, AM Chu, L Ni, C Connelly, L Riles, S Véronneau, S Dow, A Lucau-Danila, K Anderson, B André, AP Arkin, A Astromoff, M El-Bakkoury, R Bangham, R Benito, S Brachat, S Campanaro, M Curtiss, K Davis, A Deutschbauer, KD Entian, P Flaherty, F Foury, DJ Garfinkel, M Gerstein, D Gotte, U Güldener, JH Hegemann, S Hempel, Z Herman, DF Jaramillo, DE Kelly, SL Kelly, P Kötter, D LaBonte, DC Lamb, N Lan, H Liang, H Liao, L Liu, C Luo, M Lussier, R Mao, P Menard, SL Ooi, JL Revuelta, CJ Roberts, M Rose, P Ross-Macdonald, B Scherens, G Schimmack, B Shafer, DD Shoemaker, S Sookhai-Mahadeo, RK Storms, JN Strathern, G Valle, M Voet, G Volckaert, CY Wang, TR Ward, J Wilhelmy, EA Winzeler, Y Yang, G Yen, E Youngman, K Yu, H Bussey, JD Boeke, M Snyder, P Philippsen, RW Davis, M Johnston (2002). "Functional profiling of the *Saccharomyces cerevisiae* genome." *Nature* 418: 387-91.
- S Balasubramanian, P Harrison, H Hegyi, P Bertone, N Luscombe, N Echols, P McGarvey, Z Zhang, M Gerstein (2002). "SNPs on human chromosomes 21 and 22 -- analysis in terms of protein features and pseudogenes." *Pharmacogenomics* 3: 393-402.
- N Echols, P Harrison, S Balasubramanian, NM Luscombe, P Bertone, Z Zhang, M Gerstein (2002). "Comprehensive analysis of amino acid and nucleotide composition in eukaryotic genomes, comparing genes and pseudogenes." *Nucleic Acids Res* 30: 2515-23.
- CE Horak, MC Mahajan, NM Luscombe, M Gerstein, SM Weissman, M Snyder (2002). "GATA-1 binding sites mapped in the beta-globin locus by using mammalian chIp-chip analysis." *Proc Natl Acad Sci U S A* 99: 2924-9.
- Y Liu, NM Luscombe, V Alexandrov, P Bertone, P Harrison, Z Zhang, M Gerstein (2002). "Structural genomics: a new era for pharmaceutical research." *Genome Biol* 3: REPORTS4004.
- A Kumar, PM Harrison, KH Cheung, N Lan, N Echols, P Bertone, P Miller, MB Gerstein, M Snyder (2002). "An integrated approach for finding overlooked genes in yeast." *Nat Biotechnol* 20: 58-63.

-- 2001 --

- J Qian, NM Luscombe, M Gerstein (2001). "Protein family and fold occurrence in genomes: power-law behaviour and evolutionary model." *J Mol Biol* 313: 673-81.
- J Qian, M Dolled-Filhart, J Lin, H Yu, M Gerstein (2001). "Beyond synexpression relationships: local clustering of time-shifted and inverted gene expression profiles identifies new, biologically relevant interactions." *J Mol Biol* 314: 1053-66.
- H Hegyi, M Gerstein (2001). "Annotation transfer for genomics: measuring functional divergence in multi-domain proteins." *Genome Res* 11: 1632-40.
- NM Luscombe, D Greenbaum, M Gerstein (2001). "What is bioinformatics? A proposed definition and overview of the field." *Methods Inf Med* 40: 346-58.
- R Das, J Junker, D Greenbaum, MB Gerstein (2001). "Global perspectives on proteins: comparing genomes in terms of folds, pathways and beyond." *Pharmacogenomics J* 1: 115-25.
- H Zhu, M Bilgin, R Bangham, D Hall, A Casamayor, P Bertone, N Lan, R Jansen, S Bidlingmaier, T Houfek, T Mitchell, P Miller, RA Dean, M Gerstein, M Snyder (2001). "Global analysis of protein activities using proteome chips." *Science* 293: 2101-5.
- V Alexandrov, M Gerstein (2001). "Calculating populations of subcellular compartments using density matrix formalism" *International Journal of Quantum Chemistry* 85: 693-696.
- WK Olson, M Bansal, SK Burley, RE Dickerson, M Gerstein, SC Harvey, U Heinemann, XJ Lu, S Neidle, Z Shakked, H Sklenar, M Suzuki, CS Tung, E Westhof, C Wolberger, HM Berman (2001). "A standard reference frame for the description of nucleic acid base-pair geometry." *J Mol Biol* 313: 229-37.

- J Tsai, N Voss, M Gerstein (2001). "Determining the minimum number of types necessary to represent the sizes of protein atoms." *Bioinformatics* 17: 949-56.
- D Greenbaum, NM Luscombe, R Jansen, J Qian, M Gerstein (2001). "Interrelating different types of genomic data, from proteome to secretome: 'oming in on function.'" *Genome Res* 11: 1463-8.
- M Gerstein, F M Richards (2001). "Protein Geometry: Distances, Areas, and Volumes" *International Tables for Crystallography* : (Volume F, Chapter 22.1.1, pages 531-539; M Rossmann & E Arnold, editors; Dordrecht: Kluwer).
- P Harrison, M Gerstein (2001). " A Bauhaus for Biologists: An Introduction to Protein Architecture by A. M. Lesk" *Trends Biochem Sci* 26: 204-205.
- P Bertone, M Gerstein (2001). "Integrative data mining: the new direction in bioinformatics." *IEEE Eng Med Biol Mag* 20: 33-40.
- Z Lian, L Wang, S Yamaga, W Bonds, Y Beazer-Barclay, Y Kluger, M Gerstein, PE Newburger, N Berliner, SM Weissman (2001). "Genomic and proteomic analysis of the myeloid differentiation program." *Blood* 98: 513-24.
- YV Subrahmanyam, S Yamaga, Y Prashar, HH Lee, NP Hoe, Y Kluger, M Gerstein, JD Goguen, PE Newburger, SM Weissman (2001). "RNA expression patterns change dramatically in human neutrophils exposed to bacteria." *Blood* 97: 2457-68.
- P Bertone, Y Kluger, N Lan, D Zheng, D Christendat, A Yee, AM Edwards, CH Arrowsmith, GT Montelione, M Gerstein (2001). "SPINE: an integrated tracking database and data mining approach for identifying feasible targets in high-throughput structural proteomics." *Nucleic Acids Res* 29: 2884-98.
- M Gerstein, B Honig (2001). "Sequences and Topology" *Current Opinion in Structural Biology* 11: 327-329.
- N Luscombe, D Greenbaum, M Gerstein (2001). "What is Bioinformatics? A Proposed Definition and Overview of the Field" *Intl. Medical Informatics Association* : (Yearbook, Pages 83-99).
- PM Harrison, N Echols, MB Gerstein (2001). "Digging for dead genes: an analysis of the characteristics of the pseudogene population in the *Caenorhabditis elegans* genome." *Nucleic Acids Res* 29: 818-30.
- J Qian, B Stenger, CA Wilson, J Lin, R Jansen, SA Teichmann, J Park, WG Krebs, H Yu, V Alexandrov, N Echols, M Gerstein (2001). "PartsList: a web-based system for dynamically ranking protein folds based on disparate attributes, including whole-genome expression and interaction information." *Nucleic Acids Res* 29: 1750-64.
- An XML Application for Genomic Data Interoperation. Cheung KH, Liu Y, Kumar K, Snyder M, Gerstein M, Miller P. IEEE International Symposium on Bio-Informatics and Biomedical Engineering (BIBE) 2001, pp. 97-103 .

-- 2000 --

- M Gerstein (2000). "Integrative database analysis in structural genomics." *Nat Struct Biol* 7 Suppl: 960-3.
- R Das, H Hegyi, M Gerstein (2000). "Genome analyses of spirochetes: a study of the protein structures, functions and metabolic pathways in *Treponema pallidum* and *Borrelia burgdorferi*." *J Mol Microbiol Biotechnol* 2: 387-92.
- D Christendat, A Yee, A Dharamsi, Y Kluger, M Gerstein, CH Arrowsmith, AM Edwards (2000). "Structural proteomics: prospects for high throughput sample preparation." *Prog Biophys Mol Biol* 73: 339-45.
- H Zhu, JF Klemic, S Chang, P Bertone, A Casamayor, KG Klemic, D Smith, M Gerstein, MA Reed, M Snyder (2000). "Analysis of yeast protein kinases using protein chips." *Nat Genet* 26: 283-9.
- A Drawid, R Jansen, M Gerstein (2000). "Genome-wide analysis relating expression level with protein subcellular localization." *Trends Genet* 16: 426-30.

- M Gerstein, R Jansen (2000). "The current excitement in bioinformatics-analysis of whole-genome expression data: how does it relate to protein structure and function?" *Curr Opin Struct Biol* 10: 574-84.
- R Das, M Gerstein (2000). "The stability of thermophilic proteins: a study based on comprehensive genome comparison." *Funct Integr Genomics* 1: 76-88.
- GJ Naylor, M Gerstein (2000). "Measuring shifts in function and evolutionary opportunity using variability profiles: a case study of the globins." *J Mol Evol* 51: 223-33.
- D Christendat, A Yee, A Dharamsi, Y Kluger, A Savchenko, JR Cort, V Booth, CD Mackereth, V Saridakis, I Ekiel, G Kozlov, KL Maxwell, N Wu, LP McIntosh, K Gehring, MA Kennedy, AR Davidson, EF Pai, M Gerstein, AM Edwards, CH Arrowsmith (2000). "Structural proteomics of an archaeon." *Nat Struct Biol* 7: 903-9.
- A Drawid, M Gerstein (2000). "A Bayesian system integrating expression data with sequence patterns for localizing proteins: comprehensive application to the yeast genome." *J Mol Biol* 301: 1059-75.
- S Balasubramanian, T Schneider, M Gerstein, L Regan (2000). "Proteomics of *Mycoplasma genitalium*: identification and characterization of unannotated and atypical proteins in a small model genome." *Nucleic Acids Res* 28: 3075-82.
- M Gerstein, J Lin, H Hegyi (2000). "Protein folds in the worm genome." *Pac Symp Biocomput* : 30-41.
- M Gerstein (2000). "Annotation of the human genome." *Science* 288: 1590.
- J Lin, M Gerstein (2000). "Whole-genome trees based on the occurrence of folds and orthologs: implications for comparing genomes on different levels." *Genome Res* 10: 808-18.
- WG Krebs, M Gerstein (2000). "The morph server: a standardized system for analyzing and visualizing macromolecular motions in a database framework." *Nucleic Acids Res* 28: 1665-75.
- A Senes, M Gerstein, DM Engelman (2000). "Statistical analysis of amino acid patterns in transmembrane helices: the GxxxG motif occurs frequently and in association with beta-branched residues at neighboring positions." *J Mol Biol* 296: 921-36.
- CA Wilson, J Kreychman, M Gerstein (2000). "Assessing annotation transfer for genomics: quantifying the relations between protein sequence, structure and function through traditional and probabilistic scores." *J Mol Biol* 297: 233-49.
- R Jansen, M Gerstein (2000). "Analysis of the yeast transcriptome with structural and functional categories: characterizing highly expressed proteins." *Nucleic Acids Res* 28: 1481-8.

-- 1999 --

- P Ross-Macdonald, PS Coelho, T Roemer, S Agarwal, A Kumar, R Jansen, KH Cheung, A Sheehan, D Symoniatis, L Umansky, M Heidtman, FK Nelson, H Iwasaki, K Hager, M Gerstein, P Miller, GS Roeder, M Snyder (1999). "Large-scale analysis of the yeast genome by transposon tagging and gene disruption." *Nature* 402: 413-8.
- M Gerstein, C Chothia (1999). "Perspectives: signal transduction. Proteins in motion." *Science* 285: 1682-3.
- M Gerstein, R Jansen, T Johnson, J Tsai, W Krebs (1999). "Motions in a Database Framework: from Structure to Sequence" *Rigidity Theory and Applications* : 401-442 (ed. M F Thorpe and P M Duxbury, Kluwer Academic/Plenum Publishers).
- M Gerstein (1999). "E-publishing on the Web: promises, pitfalls, and payoffs for bioinformatics." *Bioinformatics* 15: 429-31.
- ES Brodtkin, M Gerstein (1999). "'E-biomed' and clinical research." *N Engl J Med* 341: 1080; author reply 1081.

- J Tsai, R Taylor, C Chothia, M Gerstein (1999). "The packing density in proteins: standard radii and volumes." *J Mol Biol* 290: 253-66.
- SA Teichmann, C Chothia, M Gerstein (1999). "Advances in structural genomics." *Curr Opin Struct Biol* 9: 390-9.
- M Gerstein (1999). "Building the future of biocomputing." *Nature* 399: 101.
- H Hegyi, M Gerstein (1999). "The relationship between protein structure and function: a comprehensive survey with application to the yeast genome." *J Mol Biol* 288: 147-64.
- M Gerstein (1999). "Forging links in an electronic paper chain." *Nature* 398: 20.

-- 1998 --

- M Gerstein, H Hegyi (1998). "Comparing genomes in terms of protein structure: surveys of a finite parts list." *FEMS Microbiol Rev* 22: 277-304.
- M Gerstein (1998). "How representative are the known structures of the proteins in a complete genome? A comprehensive structural census." *Fold Des* 3: 497-512.
- M Gerstein (1998). "Patterns of protein-fold usage in eight microbial genomes: a comprehensive structural census." *Proteins* 33: 518-34.
- M Gerstein, M Levitt (1998). "Simulating water and the molecules of life." *Sci Am* 279: 100-5.
- M Gerstein (1998). "Measurement of the effectiveness of transitive sequence comparison, through a third 'intermediate' sequence." *Bioinformatics* 14: 707-14.
- M Gerstein, W Krebs (1998). "A database of macromolecular motions." *Nucleic Acids Res* 26: 4280-90.
- J Fu, M Gerstein, PR David, AL Gnatt, DA Bushnell, AM Edwards, RD Kornberg (1998). "Repeated tertiary fold of RNA polymerase II and implications for DNA binding." *J Mol Biol* 280: 317-22.
- M Levitt, M Gerstein (1998). "A unified statistical framework for sequence comparison and structure comparison." *Proc Natl Acad Sci U S A* 95: 5913-20.
- M Gerstein, M Levitt (1998). "Comprehensive assessment of automatic structural alignment against a manual standard, the scop classification of proteins." *Protein Sci* 7: 445-56.

-- 1997 --

- M Gerstein (1997). "A structural census of genomes: comparing bacterial, eukaryotic, and archaeal genomes in terms of protein structure." *J Mol Biol* 274: 562-76.
- M Gerstein, M Levitt (1997). "A structural census of the current population of protein sequences." *Proc Natl Acad Sci U S A* 94: 11911-6.
- C Chothia, M Gerstein (1997). "Protein evolution. How far can sequences diverge?" *Nature* 385: 579, 581.
- M Levitt, M Gerstein, E Huang, S Subbiah, J Tsai (1997). "Protein folding: the endgame." *Annu Rev Biochem* 66: 549-79.
- R Schmidt, M Gerstein, RB Altman (1997). "LPFC: an Internet library of protein family core structures." *Protein Sci* 6: 246-8.

-- 1996 --

- R Becker, M Gerstein (1996). Practicing Cyberlaw in the Year 2000. *New Jersey Lawyer Magazine* 179: 12-27 (September).
- J Tsai, M Gerstein, M Levitt (1996). Keeping the Shape but Changing the Charges: A Simulation Study of Urea and its Isosteric Analogues. *Journal of Chemical Physics* 104: 9417-9430.

M Gerstein, C Chothia (1996). "Packing at the protein-water interface." *Proc Natl Acad Sci U S A* 93: 10167-72.

M Gerstein, M Levitt (1996). "Using iterative dynamic programming to obtain accurate pairwise and multiple alignments of protein structures." *Proc Int Conf Intell Syst Mol Biol* 4: 59-67.

-- 1995 --

M Gerstein, RB Altman (1995). "Using a measure of structural variation to define a core for the globins." *Comput Appl Biosci* 11: 633-44.

M Suzuki, M Gerstein (1995). "Binding geometry of alpha-helices that recognize DNA." *Proteins* 23: 525-35.

M Gerstein, RB Altman (1995). "Average core structures and variability measures for protein families: application to the immunoglobulins." *J Mol Biol* 251: 161-75.

M Gerstein, J Tsai, M Levitt (1995). "The volume of atoms on the protein surface: calculated from simulation, using Voronoi polyhedra." *J Mol Biol* 249: 955-66.

RB Altman, C Hughes, MB Gerstein (1995). "Methods for displaying macromolecular structural uncertainty: application to the globins." *J Mol Graph* 13: 142-52, 109-2.

M Suzuki, N Yagi, M Gerstein (1995). "DNA recognition and superstructure formation by helix-turn-helix proteins." *Protein Eng* 8: 329-38.

M Suzuki, SE Brenner, M Gerstein, N Yagi (1995). "DNA recognition code of transcription factors." *Protein Eng* 8: 319-28.

-- 1994 --

Purcell's early work on NMR: Contingency versus Inevitability . M Gerstein (1994) *American Journal of Physics* 62: 596-601.

M Suzuki, M Gerstein, N Yagi (1994). "Stereochemical basis of DNA recognition by Zn fingers." *Nucleic Acids Res* 22: 3397-405.

Y Harpaz, M Gerstein, C Chothia (1994). "Volume changes on protein folding." *Structure* 2: 641-9.

M Gerstein, AM Lesk, C Chothia (1994). "Structural mechanisms for domain movements in proteins." *Biochemistry* 33: 6739-49.

M Suzuki, D Neuhaus, M Gerstein, S Aimoto (1994). "Solution structure of the DNA binding octapeptide repeat of the K10 gene product." *Protein Eng* 7: 461-70.

M Gerstein, EL Sonnhammer, C Chothia (1994). "Volume changes in protein evolution." *J Mol Biol* 236: 1067-78.

RB Altman, M Gerstein (1994). "Finding an average core structure: application to the globins." *Proc Int Conf Intell Syst Mol Biol* 2: 19-27.

-- 1993 --

Simulation of Water around a Model Protein Helix. 1. Two-dimensional Projections of Solvent Structure. M Gerstein, R Lynden-Bell (1993) *Journal of Physical Chemistry* 97: 2982-2991. Simulation of Water around a Model Protein Helix. 2. The Relative Contributions of Packing, Hydrophobicity, and Hydrogen Bonding. M Gerstein, R Lynden-Bell (1993) *Journal of Physical Chemistry* 97: 2991-2999..

M Gerstein, BF Anderson, GE Norris, EN Baker, AM Lesk, C Chothia (1993). "Domain closure in lactoferrin. Two hinges produce a see-saw motion between alternative close-packed interfaces." *J Mol Biol* 234: 357-72.

M Suzuki, M Gerstein, T Johnson (1993). "An NMR study on the DNA-binding SPKK motif and a model for its interaction with DNA." *Protein Eng* 6: 565-74.

M Gerstein, RM Lynden-Bell (1993). "What is the natural boundary of a protein in solution?" *J Mol Biol* 230: 641-50.

M Gerstein, G Schulz, C Chothia (1993). "Domain closure in adenylate kinase. Joints on either side of two helices close like neighboring fingers." *J Mol Biol* 229: 494-501.

S Subramaniam, M Gerstein, D Oesterhelt, R Henderson (1993). "Electron diffraction analysis of structural changes in the photocycle of bacteriorhodopsin." *EMBO J* 12: 1-8.

-- 1992 --

A Resolution-Sensitive Procedure for Comparing Protein Surfaces and its Application to the Comparison of Antigen-Combining Sites. M Gerstein (1992) *Acta Crystallographica* A48: 271-276.

I De Baere, L Liu, L Moens, J Van Beeumen, C Gielens, J Richelle, C Trotman, J Finch, M Gerstein, M Perutz (1992). "Polar zipper sequence in the high-affinity hemoglobin of *Ascaris suum*: amino acid sequence and structural interpretation." *Proc Natl Acad Sci U S A* 89: 4638-42.

-- 1991 --

M Gerstein, C Chothia (1991). "Analysis of protein loop closure. Two types of hinges produce one motion in lactate dehydrogenase." *J Mol Biol* 220: 133-49.

-- 1987 --

Inverse Problem for Synchrotron Radiation in the Presence of Noise. N Fisch, A Kritz, M Gerstein (1987) *Proceedings of the Sixth Joint Workshop on Electron Cyclotron Emission and Electron Cyclotron Resonance Heating*. (eds. A Riviere, A Costley), 23-30 (Oxford, 16-17 September).

Notes on Scientific Publications

(As of 31 Dec. 2019)

a) In total, 580 scientific publication.

Not including in press or submitted articles or popular pieces, such as 12 published newspaper Op-Eds. This list does include a few significant CS-oriented conference papers but other than that mostly excludes conference proceedings and workshops.

b) H-index for M Gerstein is 166

(according to Google Scholar, scholar.google.com/citations?user=YvjuUugAAAAJ)

c) Highly Cited Researchers (HCR) list inclusion.

On Thomson Reuters list, in 2014 to 2016. On succeeding Clarivate list, in 2017 to 2019.

d) Senior Author Status.

In the publication list, if M Gerstein is not a last or first author, he is **not** considered to be a "corresponding" or "senior" author except as noted by the asterisks (*) in the list below:

Li... PsychENCODE Consortium... Sanders, Sullivan,

Gerstein*, Lein*, Knowles*, Sestan* ('18). *Science* 362: 1264

Carlyle... Sestan, **Gerstein***, Nairn* ('18). *J Proteome Res* 17: 3431

Sudmant... 1000 Genomes Project, Mills*, **Gerstein***, Bashir*, Stegle*, Devine*, Lee*, Eichler*, Korbelt* ('15). *Nature* 526: 75

Abyzov... Urban*, **Gerstein***, Vaccarino* ('12) *Nature* 492: 438

Gianoulis... **Gerstein***, Strobel* ('12). *PLoS Genet* 8: e1002558

MacArthur... 1000 Genomes Project... **Gerstein***, Tyler-Smith* ('12). *Science* 335: 823

Mills... Eichler*, **Gerstein***, Hurles*, Lee*, McCarroll*, Korbelt*, 1000 Genomes Project ('11). *Nature* 470: 59

Bertone... **Gerstein***, Snyder* ('04). *Science* 306: 2242

Other Writings & Presentations (as of 30 Sep. 2019)

Opinion Pieces

- D Greenbaum & M Gerstein (2008). "Danger: Sharing Gene Data", Hartford Courant, July 10, pg. A11 (Op-ed)
- D Greenbaum & M Gerstein (2008). "Personal genomics requires redefining privacy -- The human blueprint: dangerous secrets", SF Chronicle, Nov. 2, Page 2 (Insight)
- M Seringhaus & M Gerstein (2009). "Putting too much information online can erode individual privacy", Hartford Courant, June 5 (Op-ed)
- D Greenbaum & M Gerstein (2010). "Exploring genetics of professional athletes", SF Chronicle, May 2, Page E-4 (Insight)
- D Greenbaum & M Gerstein (2012). "The Age of Genetically Optimized Sports", Wall Street Journal, July 24, Page A13 (Opinion)
- D Greenbaum & M Gerstein (2013). "Your DNA vulnerable to snooping, too?", USA Today, July 27 (Opinion)
- D Greenbaum & M Gerstein (2013). "Proceed with Caution," The Scientist, Oct 1
- D Greenbaum & M Gerstein (2015). "Too big to close down: Websites need regulation like utilities", SF Chronicle, April 24 (Opinion)
- D Greenbaum & M Gerstein (2015). "Why can employers fingerprint, but not test workers' DNA?", SF Chronicle, July 10 (Opinion)
- D Greenbaum & M Gerstein (2016). "Going beyond geek chic -- CeBIT", SF Chronicle, March 10 (Opinion)
- D Greenbaum & M Gerstein (2017). "Pooling data from wearables could boost health benefits", SF Chronicle, Feb. 24 (Opinion)
- D Greenbaum & M Gerstein (2018). "Our smart devices must be made to forget us", SF Chronicle, Oct. 18 (Opinion)

Recorded Panel Discussions & Interviews

- M Gerstein (2008). "A Great Historical Document - The Human Genome", Futures in Biotech 34 (podcast moderated by M Pelletier)
- "A Closer Look at Personal Genomic Testing", Inforum Genomics Panel, at the Commonwealth Club of California, including L Avey, D Ballon, D Magnus, M Gerstein, J Rae-Dupree (2009)
- "Whose DNA is it?", a panel discussion on Personal Genomics, on the Agenda with Steve Paikin, as part of the Quantum to Cosmos Festival (Q2C) in Waterloo, ON, 21 Oct. 2009
- "Genomics, Proteomics, Cellular Immunity, and Anti-Matter", a panel discussion moderated by M Pelletier, including V Racaniello, A Nantel, M Gerstein, and G Farr. Futures in Biotech 71 (22 Nov. 2010)
- M Gerstein (2011). "Bioinformatics: Essential Gene names Skewed in a Network of Blame", Futures in Biotech 83 (podcast moderated by M Pelletier)
- "6 PhDs Piled High And Deep", a panel discussion moderated by M Pelletier, including G Farr, D Thomas, M Gerstein, S Melov, and J Sanchez. Futures in Biotech 91 (16 Dec. 2011)
- M Gerstein (2014). "What in the World", Sirius XM Radio Canada, 60' on 20 Nov. (Interview by Richard Garner)
- M Gerstein (2015). "What Now? Going Beyond the \$1,000 Genome", Mendelspod, 17 Sept. (podcast moderated by T Timpson)