**Gerstein lab experience in brain and cancer genomics**

**I. Lab experience in data fusion for brain and cancer genomics studies**

We have extensive experience with the ENCODE(1-3) modENCODE(4, 5), 1,000 Genomes (6-8), Brainspan, and PsychENCODE(9) consortia, in which we served a variety of leadership roles (e.g.., co-lead of the analysis working group for ENCODE and modENCODE)(1, 2, 5, 10). We recently published a full integrative analysis of data from the PsychENCODE project, which includes large-scale genomic and epigenetic data (see below) for multiple brain diseases and enabled significantly improved predictions of disease from genotype(11). We also have extensive experience analyzing cancer genomes through our participation in TCGA and PCAWG consortia. We participated in the TCGA consortium studies of prostate (12) and kidney (13) cancers and conducted a targeted study of non-coding mutations in TCGA kidney papillary cancer(14). We have also developed tools for somatic variant calling(15). Currently, we are co-leading a PCAWG sub-group to investigate the impact of non-coding mutations.

We have analyzed patterns of variation in non-coding regions and their coding targets (2, 16, 17). In recent projects (7, 18), we integrated multiple methods into a comprehensive prioritization pipeline called FunSeq. The pipeline identifies sensitive regions, e.g. TF binding sites and enhancers, with annotations under high selective pressure, links non-coding mutations to their target genes, prioritizes mutations based on network connectivity. Using integrated data from large-scale resources (including ENCODE and the 1,000 Genomes Project) with cancer genomics data, FunSeq can prioritize known TERT promoter driver mutations. To better annotate noncoding regions, we have developed matchedFilter, a signal processing based enhancer predictor that uses shape-matching filters of histone modification signals inputs and enhancer reporter assays for model training(19). We also recently developed RADAR, a tool that prioritizes mutations in noncoding RNA regions, primarily using information about RNA-protein interactions(20). Further, we have developed a predictor of upstream open reading frames (uORFs), which are noncoding RNA elements that can regulate translation.

**II. Lab experience in statistical modeling for genomic interpretation**

Our lab has expertise in modelling genomic variations and their impact on regulatory networks, and specifically to detect driver mutations. Cancer driver genes are expected to have enriched levels of mutations, but different genome regions have different background mutation rates. We developed a computational framework called LARVA, which employs non-coding functional elements as covariates, modeling the mutation counts of the elements with a beta-binomial distribution to handle overdispersion(21). Regional genomic features such as replication timing help better estimate local mutation rates and find mutational hotspots. Applying LARVA to 760 whole-genome tumor sequences shows that it identifies well-known non-coding drivers, such as mutations in the TERT promoter, in addition to uncovering new potential non-coding driver regions. We also created a nonparametric tool, called Mutations Overburdening Annotations Tool (MOAT), which uses permutation-based tests to search for overburdening of epigenetic annotations, e.g. transcription factor binding sites(22).

We have also used TCGA RNA-seq data to develop and apply multiple tools. For example, we developed a computational method called DREISS for analyzing the “Dynamics of gene expression driven by Regulatory networks, both External and Internal, based on State Space models”(23). DREISS employs dimensionality reduction to help identify canonical temporal dynamics (e.g., degradation, growth, and oscillation) representing the regulatory effects emanating from various subsystems. Another such tool that we developed, Loregic, computationally integrates gene expression and regulatory network data to characterize the cooperativity of regulatory factors(24). These tools can be applied to cancer samples to identify crucial regulatory motifs.

Further, as part of the PsychENCODE consoritum, we developed an interpretable deep-learning framework, the Deep Structured Phenotype Network (DSPN), to model the full interaction between genotype and phenotype ((11, 25). DSPN connects molecular level networks–including enhancer-gene, TF-gene, QTL, and other regulatory linkages–to phenotypes to diseases using interpretable deep-learning model. Our model uses a conditional Deep Boltzmann Machine architecture and introduces lateral connectivity at the visible layer to embed the biological structure learned from the regulatory network and QTL linkages as sparse connections between nodes. Further, a rank-statistic based interpretation scheme which allows us to functionally annotate hidden nodes and prioritize them relative to disorders, generating a hierarchy of ‘higher-order modules’ (generalizing gene co-expression modules) linked to traits of interest. Our model improves disease prediction (by 6-fold compared to additive polygenic risk scores), highlights key genes for disorders, and allows imputation of missing transcriptome information from genotype data alone.

**III. Lab experience in structural modeling**

We have developed a number of tools to evaluate mutation deleteriousness in the context of protein structures and their interactions. Our comprehensive framework incorporates protein structure and dynamics for predicting allosteric residues both on the surface and in the interior (stress.molmovdb.org). Our STRESS tool identifies mutations that might affect allosteric hotspots in proteins, which can be key to protein function(26). This has been added to our suite of tools in MolMovDB, in which users visualize conformational changes(27). Using RigidFinder, one may predict regions that remain relatively rigid upon such large-scale conformational changes(28). In addition, DynaSIN (which is also a part of the MolMovDB suite) catalogues instances in which protein conformational changes contribute favorably to protein interactions. Localized structural frustration identifies key functional regions(29). In addition, our Intensification tool searches for deleterious mutations particularly within repeated regions of proteins (30). GenoDock is our recently published tool to evaluate and predict the effects of nonsynonymous variants on protein-ligands interactions (genodock.molmovdb.org). In addition, HotCommics models conformational change to identify mutational clusters within protein structures that harbor significantly high numbers of cancer-associated variants. Using the output of our HotCommics framework, ontological and pathway enrichment analyses have implicated genes with predicted hotspots to be enriched in biological processes associated with tumor progression.

1. Encode Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489(7414):57-74. doi: 10.1038/nature11247. PubMed PMID: 22955616; PMCID: PMC3439153.

2. Gerstein MB, Kundaje A, Hariharan M, Landt SG, Yan KK, Cheng C, Mu XJ, Khurana E, Rozowsky J, Alexander R, Min R, Alves P, Abyzov A, Addleman N, Bhardwaj N, Boyle AP, Cayting P, Charos A, Chen DZ, Cheng Y, Clarke D, Eastman C, Euskirchen G, Frietze S, Fu Y, Gertz J, Grubert F, Harmanci A, Jain P, Kasowski M, Lacroute P, Leng JJ, Lian J, Monahan H, O'Geen H, Ouyang Z, Partridge EC, Patacsil D, Pauli F, Raha D, Ramirez L, Reddy TE, Reed B, Shi M, Slifer T, Wang J, Wu L, Yang X, Yip KY, Zilberman-Schapira G, Batzoglou S, Sidow A, Farnham PJ, Myers RM, Weissman SM, Snyder M. Architecture of the human regulatory network derived from ENCODE data. Nature. 2012;489(7414):91-100. Epub 2012/09/08. doi: 10.1038/nature11245. PubMed PMID: 22955619; PMCID: PMC4154057.

3. Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, Schlesinger F, Xue C, Marinov GK, Khatun J, Williams BA, Zaleski C, Rozowsky J, Roder M, Kokocinski F, Abdelhamid RF, Alioto T, Antoshechkin I, Baer MT, Bar NS, Batut P, Bell K, Bell I, Chakrabortty S, Chen X, Chrast J, Curado J, Derrien T, Drenkow J, Dumais E, Dumais J, Duttagupta R, Falconnet E, Fastuca M, Fejes-Toth K, Ferreira P, Foissac S, Fullwood MJ, Gao H, Gonzalez D, Gordon A, Gunawardena H, Howald C, Jha S, Johnson R, Kapranov P, King B, Kingswood C, Luo OJ, Park E, Persaud K, Preall JB, Ribeca P, Risk B, Robyr D, Sammeth M, Schaffer L, See LH, Shahab A, Skancke J, Suzuki AM, Takahashi H, Tilgner H, Trout D, Walters N, Wang H, Wrobel J, Yu Y, Ruan X, Hayashizaki Y, Harrow J, Gerstein M, Hubbard T, Reymond A, Antonarakis SE, Hannon G, Giddings MC, Ruan Y, Wold B, Carninci P, Guigo R, Gingeras TR. Landscape of transcription in human cells. Nature. 2012;489(7414):101-8. Epub 2012/09/08. doi: 10.1038/nature11233. PubMed PMID: 22955620; PMCID: PMC3684276.

4. Gerstein MB, Rozowsky J, Yan KK, Wang D, Cheng C, Brown JB, Davis CA, Hillier L, Sisu C, Li JJ, Pei B, Harmanci AO, Duff MO, Djebali S, Alexander RP, Alver BH, Auerbach R, Bell K, Bickel PJ, Boeck ME, Boley NP, Booth BW, Cherbas L, Cherbas P, Di C, Dobin A, Drenkow J, Ewing B, Fang G, Fastuca M, Feingold EA, Frankish A, Gao G, Good PJ, Guigo R, Hammonds A, Harrow J, Hoskins RA, Howald C, Hu L, Huang H, Hubbard TJ, Huynh C, Jha S, Kasper D, Kato M, Kaufman TC, Kitchen RR, Ladewig E, Lagarde J, Lai E, Leng J, Lu Z, MacCoss M, May G, McWhirter R, Merrihew G, Miller DM, Mortazavi A, Murad R, Oliver B, Olson S, Park PJ, Pazin MJ, Perrimon N, Pervouchine D, Reinke V, Reymond A, Robinson G, Samsonova A, Saunders GI, Schlesinger F, Sethi A, Slack FJ, Spencer WC, Stoiber MH, Strasbourger P, Tanzer A, Thompson OA, Wan KH, Wang G, Wang H, Watkins KL, Wen J, Wen K, Xue C, Yang L, Yip K, Zaleski C, Zhang Y, Zheng H, Brenner SE, Graveley BR, Celniker SE, Gingeras TR, Waterston R. Comparative analysis of the transcriptome across distant species. Nature. 2014;512(7515):445-8. Epub 2014/08/29. doi: 10.1038/nature13424. PubMed PMID: 25164755; PMCID: PMC4155737.

5. Gerstein MB, Lu ZJ, Van Nostrand EL, Cheng C, Arshinoff BI, Liu T, Yip KY, Robilotto R, Rechtsteiner A, Ikegami K, Alves P, Chateigner A, Perry M, Morris M, Auerbach RK, Feng X, Leng J, Vielle A, Niu W, Rhrissorrakrai K, Agarwal A, Alexander RP, Barber G, Brdlik CM, Brennan J, Brouillet JJ, Carr A, Cheung MS, Clawson H, Contrino S, Dannenberg LO, Dernburg AF, Desai A, Dick L, Dose AC, Du J, Egelhofer T, Ercan S, Euskirchen G, Ewing B, Feingold EA, Gassmann R, Good PJ, Green P, Gullier F, Gutwein M, Guyer MS, Habegger L, Han T, Henikoff JG, Henz SR, Hinrichs A, Holster H, Hyman T, Iniguez AL, Janette J, Jensen M, Kato M, Kent WJ, Kephart E, Khivansara V, Khurana E, Kim JK, Kolasinska-Zwierz P, Lai EC, Latorre I, Leahey A, Lewis S, Lloyd P, Lochovsky L, Lowdon RF, Lubling Y, Lyne R, MacCoss M, Mackowiak SD, Mangone M, McKay S, Mecenas D, Merrihew G, Miller DM, 3rd, Muroyama A, Murray JI, Ooi SL, Pham H, Phippen T, Preston EA, Rajewsky N, Ratsch G, Rosenbaum H, Rozowsky J, Rutherford K, Ruzanov P, Sarov M, Sasidharan R, Sboner A, Scheid P, Segal E, Shin H, Shou C, Slack FJ, Slightam C, Smith R, Spencer WC, Stinson EO, Taing S, Takasaki T, Vafeados D, Voronina K, Wang G, Washington NL, Whittle CM, Wu B, Yan KK, Zeller G, Zha Z, Zhong M, Zhou X, mod EC, Ahringer J, Strome S, Gunsalus KC, Micklem G, Liu XS, Reinke V, Kim SK, Hillier LW, Henikoff S, Piano F, Snyder M, Stein L, Lieb JD, Waterston RH. Integrative analysis of the Caenorhabditis elegans genome by the modENCODE project. Science. 2010;330(6012):1775-87. Epub 2010/12/24. doi: 10.1126/science.1196914. PubMed PMID: 21177976; PMCID: PMC3142569.

6. Genomes Project C, Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, Hurles ME, McVean GA. A map of human genome variation from population-scale sequencing. Nature. 2010;467(7319):1061-73. Epub 2010/10/29. doi: 10.1038/nature09534. PubMed PMID: 20981092; PMCID: PMC3042601.

7. Khurana E, Fu Y, Colonna V, Mu XJ, Kang HM, Lappalainen T, Sboner A, Lochovsky L, Chen J, Harmanci A, Das J, Abyzov A, Balasubramanian S, Beal K, Chakravarty D, Challis D, Chen Y, Clarke D, Clarke L, Cunningham F, Evani US, Flicek P, Fragoza R, Garrison E, Gibbs R, Gumus ZH, Herrero J, Kitabayashi N, Kong Y, Lage K, Liluashvili V, Lipkin SM, MacArthur DG, Marth G, Muzny D, Pers TH, Ritchie GRS, Rosenfeld JA, Sisu C, Wei X, Wilson M, Xue Y, Yu F, Genomes Project C, Dermitzakis ET, Yu H, Rubin MA, Tyler-Smith C, Gerstein M. Integrative annotation of variants from 1092 humans: application to cancer genomics. Science. 2013;342(6154):1235587. Epub 2013/10/05. doi: 10.1126/science.1235587. PubMed PMID: 24092746; PMCID: PMC3947637.

8. Sudmant PH, Rausch T, Gardner EJ, Handsaker RE, Abyzov A, Huddleston J, Zhang Y, Ye K, Jun G, Fritz MH, Konkel MK, Malhotra A, Stutz AM, Shi X, Casale FP, Chen J, Hormozdiari F, Dayama G, Chen K, Malig M, Chaisson MJP, Walter K, Meiers S, Kashin S, Garrison E, Auton A, Lam HYK, Mu XJ, Alkan C, Antaki D, Bae T, Cerveira E, Chines P, Chong Z, Clarke L, Dal E, Ding L, Emery S, Fan X, Gujral M, Kahveci F, Kidd JM, Kong Y, Lameijer EW, McCarthy S, Flicek P, Gibbs RA, Marth G, Mason CE, Menelaou A, Muzny DM, Nelson BJ, Noor A, Parrish NF, Pendleton M, Quitadamo A, Raeder B, Schadt EE, Romanovitch M, Schlattl A, Sebra R, Shabalin AA, Untergasser A, Walker JA, Wang M, Yu F, Zhang C, Zhang J, Zheng-Bradley X, Zhou W, Zichner T, Sebat J, Batzer MA, McCarroll SA, Genomes Project C, Mills RE, Gerstein MB, Bashir A, Stegle O, Devine SE, Lee C, Eichler EE, Korbel JO. An integrated map of structural variation in 2,504 human genomes. Nature. 2015;526(7571):75-81. Epub 2015/10/04. doi: 10.1038/nature15394. PubMed PMID: 26432246; PMCID: PMC4617611.

9. Psych EC, Akbarian S, Liu C, Knowles JA, Vaccarino FM, Farnham PJ, Crawford GE, Jaffe AE, Pinto D, Dracheva S, Geschwind DH, Mill J, Nairn AC, Abyzov A, Pochareddy S, Prabhakar S, Weissman S, Sullivan PF, State MW, Weng Z, Peters MA, White KP, Gerstein MB, Amiri A, Armoskus C, Ashley-Koch AE, Bae T, Beckel-Mitchener A, Berman BP, Coetzee GA, Coppola G, Francoeur N, Fromer M, Gao R, Grennan K, Herstein J, Kavanagh DH, Ivanov NA, Jiang Y, Kitchen RR, Kozlenkov A, Kundakovic M, Li M, Li Z, Liu S, Mangravite LM, Mattei E, Markenscoff-Papadimitriou E, Navarro FC, North N, Omberg L, Panchision D, Parikshak N, Poschmann J, Price AJ, Purcaro M, Reddy TE, Roussos P, Schreiner S, Scuderi S, Sebra R, Shibata M, Shieh AW, Skarica M, Sun W, Swarup V, Thomas A, Tsuji J, van Bakel H, Wang D, Wang Y, Wang K, Werling DM, Willsey AJ, Witt H, Won H, Wong CC, Wray GA, Wu EY, Xu X, Yao L, Senthil G, Lehner T, Sklar P, Sestan N. The PsychENCODE project. Nat Neurosci. 2015;18(12):1707-12. Epub 2015/11/26. doi: 10.1038/nn.4156. PubMed PMID: 26605881; PMCID: PMC4675669.

10. Boyle AP, Araya CL, Brdlik C, Cayting P, Cheng C, Cheng Y, Gardner K, Hillier LW, Janette J, Jiang L, Kasper D, Kawli T, Kheradpour P, Kundaje A, Li JJ, Ma L, Niu W, Rehm EJ, Rozowsky J, Slattery M, Spokony R, Terrell R, Vafeados D, Wang D, Weisdepp P, Wu YC, Xie D, Yan KK, Feingold EA, Good PJ, Pazin MJ, Huang H, Bickel PJ, Brenner SE, Reinke V, Waterston RH, Gerstein M, White KP, Kellis M, Snyder M. Comparative analysis of regulatory information and circuits across distant species. Nature. 2014;512(7515):453-6. Epub 2014/08/29. doi: 10.1038/nature13668. PubMed PMID: 25164757; PMCID: PMC4336544.

11. Wang D, Liu S, Warrell J, Won H, Shi X, Navarro FCP, Clarke D, Gu M, Emani P, Yang YT, Xu M, Gandal MJ, Lou S, Zhang J, Park JJ, Yan C, Rhie SK, Manakongtreecheep K, Zhou H, Nathan A, Peters M, Mattei E, Fitzgerald D, Brunetti T, Moore J, Jiang Y, Girdhar K, Hoffman GE, Kalayci S, Gumus ZH, Crawford GE, Psych EC, Roussos P, Akbarian S, Jaffe AE, White KP, Weng Z, Sestan N, Geschwind DH, Knowles JA, Gerstein MB. Comprehensive functional genomic resource and integrative model for the human brain. Science. 2018;362(6420). Epub 2018/12/14. doi: 10.1126/science.aat8464. PubMed PMID: 30545857.

12. Cancer Genome Atlas Research N. The Molecular Taxonomy of Primary Prostate Cancer. Cell. 2015;163(4):1011-25. Epub 2015/11/07. doi: 10.1016/j.cell.2015.10.025. PubMed PMID: 26544944; PMCID: PMC4695400.

13. Cancer Genome Atlas Research N, Linehan WM, Spellman PT, Ricketts CJ, Creighton CJ, Fei SS, Davis C, Wheeler DA, Murray BA, Schmidt L, Vocke CD, Peto M, Al Mamun AA, Shinbrot E, Sethi A, Brooks S, Rathmell WK, Brooks AN, Hoadley KA, Robertson AG, Brooks D, Bowlby R, Sadeghi S, Shen H, Weisenberger DJ, Bootwalla M, Baylin SB, Laird PW, Cherniack AD, Saksena G, Haake S, Li J, Liang H, Lu Y, Mills GB, Akbani R, Leiserson MD, Raphael BJ, Anur P, Bottaro D, Albiges L, Barnabas N, Choueiri TK, Czerniak B, Godwin AK, Hakimi AA, Ho TH, Hsieh J, Ittmann M, Kim WY, Krishnan B, Merino MJ, Mills Shaw KR, Reuter VE, Reznik E, Shelley CS, Shuch B, Signoretti S, Srinivasan R, Tamboli P, Thomas G, Tickoo S, Burnett K, Crain D, Gardner J, Lau K, Mallery D, Morris S, Paulauskis JD, Penny RJ, Shelton C, Shelton WT, Sherman M, Thompson E, Yena P, Avedon MT, Bowen J, Gastier-Foster JM, Gerken M, Leraas KM, Lichtenberg TM, Ramirez NC, Santos T, Wise L, Zmuda E, Demchok JA, Felau I, Hutter CM, Sheth M, Sofia HJ, Tarnuzzer R, Wang Z, Yang L, Zenklusen JC, Zhang J, Ayala B, Baboud J, Chudamani S, Liu J, Lolla L, Naresh R, Pihl T, Sun Q, Wan Y, Wu Y, Ally A, Balasundaram M, Balu S, Beroukhim R, Bodenheimer T, Buhay C, Butterfield YS, Carlsen R, Carter SL, Chao H, Chuah E, Clarke A, Covington KR, Dahdouli M, Dewal N, Dhalla N, Doddapaneni HV, Drummond JA, Gabriel SB, Gibbs RA, Guin R, Hale W, Hawes A, Hayes DN, Holt RA, Hoyle AP, Jefferys SR, Jones SJ, Jones CD, Kalra D, Kovar C, Lewis L, Li J, Ma Y, Marra MA, Mayo M, Meng S, Meyerson M, Mieczkowski PA, Moore RA, Morton D, Mose LE, Mungall AJ, Muzny D, Parker JS, Perou CM, Roach J, Schein JE, Schumacher SE, Shi Y, Simons JV, Sipahimalani P, Skelly T, Soloway MG, Sougnez C, Tam A, Tan D, Thiessen N, Veluvolu U, Wang M, Wilkerson MD, Wong T, Wu J, Xi L, Zhou J, Bedford J, Chen F, Fu Y, Gerstein M, Haussler D, Kasaian K, Lai P, Ling S, Radenbaugh A, Van Den Berg D, Weinstein JN, Zhu J, Albert M, Alexopoulou I, Andersen JJ, Auman JT, Bartlett J, Bastacky S, Bergsten J, Blute ML, Boice L, Bollag RJ, Boyd J, Castle E, Chen YB, Cheville JC, Curley E, Davies B, DeVolk A, Dhir R, Dike L, Eckman J, Engel J, Harr J, Hrebinko R, Huang M, Huelsenbeck-Dill L, Iacocca M, Jacobs B, Lobis M, Maranchie JK, McMeekin S, Myers J, Nelson J, Parfitt J, Parwani A, Petrelli N, Rabeno B, Roy S, Salner AL, Slaton J, Stanton M, Thompson RH, Thorne L, Tucker K, Weinberger PM, Winemiller C, Zach LA, Zuna R. Comprehensive Molecular Characterization of Papillary Renal-Cell Carcinoma. N Engl J Med. 2016;374(2):135-45. Epub 2015/11/05. doi: 10.1056/NEJMoa1505917. PubMed PMID: 26536169; PMCID: PMC4775252.

14. Li S, Shuch BM, Gerstein MB. Whole-genome analysis of papillary kidney cancer finds significant noncoding alterations. PLoS genetics. 2017;13(3):e1006685. Epub 2017/03/31. doi: 10.1371/journal.pgen.1006685. PubMed PMID: 28358873; PMCID: PMC5391127.

15. Fang LT, Afshar PT, Chhibber A, Mohiyuddin M, Fan Y, Mu JC, Gibeling G, Barr S, Asadi NB, Gerstein MB, Koboldt DC, Wang W, Wong WH, Lam HY. An ensemble approach to accurately detect somatic mutations using SomaticSeq. Genome Biol. 2015;16:197. Epub 2015/09/19. doi: 10.1186/s13059-015-0758-2. PubMed PMID: 26381235; PMCID: PMC4574535.

16. Yip KY, Cheng C, Bhardwaj N, Brown JB, Leng J, Kundaje A, Rozowsky J, Birney E, Bickel P, Snyder M, Gerstein M. Classification of human genomic regions based on experimentally determined binding sites of more than 100 transcription-related factors. Genome Biol. 2012;13(9):R48. Epub 2012/09/07. doi: 10.1186/gb-2012-13-9-r48. PubMed PMID: 22950945; PMCID: PMC3491392.

17. Mu XJ, Lu ZJ, Kong Y, Lam HY, Gerstein MB. Analysis of genomic variation in non-coding elements using population-scale sequencing data from the 1000 Genomes Project. Nucleic Acids Res. 2011;39(16):7058-76. Epub 2011/05/21. doi: 10.1093/nar/gkr342. PubMed PMID: 21596777; PMCID: PMC3167619.

18. Fu Y, Liu Z, Lou S, Bedford J, Mu XJ, Yip KY, Khurana E, Gerstein M. FunSeq2: a framework for prioritizing noncoding regulatory variants in cancer. Genome Biol. 2014;15(10):480. Epub 2014/10/03. doi: 10.1186/s13059-014-0480-5. PubMed PMID: 25273974; PMCID: PMC4203974.

19. Sethi A, Gu M, Gumusgoz E, Chan L, Yan K-K, Rozowsky J, Barozzi I, Afzal V, Akiyama J, Plajzer-Frick I, Yan C, Pickle C, Kato M, Garvin T, Pham Q, Harrington A, Mannion B, Lee E, Fukuda-Yuzawa Y, Visel A, Dickel DE, Yip K, Sutton R, Pennacchio LA, Gerstein M. A cross-organism framework for supervised enhancer prediction with epigenetic pattern recognition and targeted validation. bioRxiv. 2018.

20. Zhang J, Liu J, Lee D, Feng J-J, Lochovsky L, Lou S, Rutenberg-Schoenberg M, Gerstein M. RADAR: annotation and prioritization of variants in the post-transcriptional regulome of RNA-binding proteins. bioRxiv. 2018.

21. Lochovsky L, Zhang J, Fu Y, Khurana E, Gerstein M. LARVA: an integrative framework for large-scale analysis of recurrent variants in noncoding annotations. Nucleic Acids Res. 2015;43(17):8123-34. Epub 2015/08/26. doi: 10.1093/nar/gkv803. PubMed PMID: 26304545; PMCID: PMC4787796.

22. Lochovsky L, Zhang J, Gerstein M. MOAT: efficient detection of highly mutated regions with the Mutations Overburdening Annotations Tool. Bioinformatics. 2018;34(6):1031-3. Epub 2017/11/10. doi: 10.1093/bioinformatics/btx700. PubMed PMID: 29121169; PMCID: PMC5860157.

23. Wang D, He F, Maslov S, Gerstein M. DREISS: Using State-Space Models to Infer the Dynamics of Gene Expression Driven by External and Internal Regulatory Networks. PLoS Comput Biol. 2016;12(10):e1005146. Epub 2016/10/21. doi: 10.1371/journal.pcbi.1005146. PubMed PMID: 27760135; PMCID: PMC5070849.

24. Li MJ, Deng J, Wang P, Yang W, Ho SL, Sham PC, Wang J, Li M. wKGGSeq: A Comprehensive Strategy-Based and Disease-Targeted Online Framework to Facilitate Exome Sequencing Studies of Inherited Disorders. Hum Mutat. 2015;36(5):496-503. doi: 10.1002/humu.22766. PubMed PMID: 25676918.

25. Warrell J, Mohsen H, Gerstein M. Rank Projection Trees for Multilevel Neural Network Interpretation. arXiv preprint arXiv:181200172. 2018.

26. Clarke D, Sethi A, Li S, Kumar S, Chang RWF, Chen J, Gerstein M. Identifying Allosteric Hotspots with Dynamics: Application to Inter- and Intra-species Conservation. Structure. 2016;24(5):826-37. Epub 2016/04/14. doi: 10.1016/j.str.2016.03.008. PubMed PMID: 27066750; PMCID: PMC4883016.

27. Echols N, Milburn D, Gerstein M. MolMovDB: analysis and visualization of conformational change and structural flexibility. Nucleic Acids Res. 2003;31(1):478-82. Epub 2003/01/10. PubMed PMID: 12520056; PMCID: PMC165551.

28. Abyzov A, Bjornson R, Felipe M, Gerstein M. RigidFinder: a fast and sensitive method to detect rigid blocks in large macromolecular complexes. Proteins. 2010;78(2):309-24. Epub 2009/08/26. doi: 10.1002/prot.22544. PubMed PMID: 19705487.

29. Kumar S, Clarke D, Gerstein M. Localized structural frustration for evaluating the impact of sequence variants. Nucleic Acids Res. 2016;44(21):10062-73. Epub 2016/12/05. doi: 10.1093/nar/gkw927. PubMed PMID: 27915290; PMCID: PMC5137452.

30. Chen J, Wang B, Regan L, Gerstein M. Intensification: A Resource for Amplifying Population-Genetic Signals with Protein Repeats. J Mol Biol. 2017;429(3):435-45. Epub 2016/12/13. doi: 10.1016/j.jmb.2016.12.003. PubMed PMID: 27939289; PMCID: PMC5420328.