1. Experience in analysis of omics data

1.1 Experience in transcriptomics

We have extensive expertise performing transcriptome analyses, developing a wide range of customized tools, and building standardized pipelines for analysis and uniform processing of both long and short RNA-sequencing (RNA-Seq) data. These tools have been evaluated and implemented in several major consortia.

For general RNA-Seq analysis, we have developed an efficient in-house data processing workflow for long RNA-Seq data that includes data organization, format conversion and quality assessment. RSEQtools (http://rseqtools.gersteinlab.org) is a computational package that enables users to quantify the expression of annotated RNAs and identify splice sites and gene models [1]. Comparisons between RNA-Seq samples, and to other genome-wide data, are facilitated by our Aggregation and Correlation Toolbox, which compares genomic signal tracks [2]. We developed incRNA [3] to predict novel non-coding RNAs (ncRNAs) using known ncRNAs of various biotypes. We created FusionSeq to detect transcripts that arise due to trans-splicing or chromosomal translocations [4]. We also constructed IQSeq [5], which calculates the relative and absolute abundance of contributing transcript isoforms to a gene from RNA-Seq data. We developed the AlleleSeq tool [6] that combines diploid genomic information with RNA-Seq data to identify transcripts showing allele-specific expression. We have further developed Pseudo-seq [7] and PseudoPipe [8] to address the issue of quantification of pseudogene expression.

We recently developed the extracellular RNA (exRNA) processing toolkit (Submitted) available at http://github.gersteinlab.org/exceRpt, which includes a set of tools and a pipeline designed for comprehensive analysis of small RNA-Seq datasets. We specifically designed exceRpt to handle technical issues that are often characteristic of small RNA-Seq samples, such as read preprocessing, filtering and alignment, biotype abundance estimation, visualization and quality assessment. The exceRpt pipeline has been used for uniform processing of hundreds of RNA-Seq datasets submitted to the exRNA Atlas (http://exrna-atlas.org) repository.

We have extensive experience conducting integrative analyses of large sets of RNA-Seq data. We have developed and analyzed multiple RNA-Seq flows in the context of large consortia, and have implemented both tools that we developed and other popular tools such as Bowtie [9] and Tophat [10].

1.2 Experience in proteomics

We have substantial experience analyzing proteomic data [11-13] and integrating it with genomic data. For example, to combine mass spectrometry (MS) proteomic and transcriptomic data [14, 15], we constructed a web tool called PARE (Protein Abundance and mRNA Expression; <http://proteomics.gersteinlab.org>) [16]. We also published the tool Empire [17], which uses transcript-level RNA-Seq expression as a prior likelihood and enables users to directly estimate protein isoform abundances from liquid chromatography MS/MS (LC–MS/MS), an approach derived from the principle that most genes appear to be expressed as a single dominant isoform in a given cell type or tissue. We have also led studies interpreting protein-protein interactions based on data from proteomic experiments [18, 19]. We have been members of numerous NIH proteomics projects and consortia, including the Northeast Structural Genomics Consortium, the National Heart, Lung and Blood Institute Proteomics Center and the Yale/National Institute on Drug Abuse Neuroproteomics Center, and have conducted analyses on the various types of large-scale proteomic data generated by these consortia [12, 20].

1.3 Experience in metabolomics

We have demonstrated that metabolic phenotypes of response to vaccination in humans revealed strong association between plasma metabolomics and PBMC transcriptomics and concluded that metabolomic phenotypes, such as inositol phosphate metabolism, influence immune outcome [21]. We also took a systems biology approach to integrate existing knowledge and novel high-throughput data that facilitates a quantitative understanding of the molecular mechanism of ω3- and ω6-PUFA metabolism in mammalian cells[22]. Additionally, the UCSD Center for Computational Biology and Bioinformatics (CCBB) collaborated on a recent metabolomics analysis of rheumatoid arthritis [23]. We will apply these proven methodologies to the identification of acute-to-chronic pain signatures and the associated mechanisms.

1.4 Experience in lipidomics

The UCSD LIPID MAPS Lipidomics Core (<http://www.ucsd-lipidmaps.org>) aims to develop the field of lipidomics, especially studies focused on targeting bioactive lipid mediators and developing biomarkers[24]. The complexity of the lipidome both in its dynamic range and structural diversity represents a major analytical challenge. To address these challenges, the LIPID MAPS Consortium is quantifying all lipid species of the mammalian lipidome. We established the first comprehensive human lipid profile in plasma and identified and quantified 600 distinct lipid molecular species across all mammalian lipid categories [25]. We profiled activated macrophages immunologically, measured over 500 discrete lipid species and mapped associated pathways, integrating transcriptomics, proteomics and lipidomics [26]. We now routinely profile various tissues of both human and animal origin for biomarker discovery and for indicators of abnormal lipid metabolism. More recently, we established lipid profiles of liver biopsy specimen and plasma from individuals with non-alcoholic fatty liver disease for biomarker development [27]. We used similar approaches to identify eicosanoid targets in various bacterial and viral infectious diseases including Lyme disease and influenza [28, 29]. Our lipidomics platform for monitoring over 200 oxidation and signal transduction consequences is the most developed platform to emerge in the metabolomics area [30-33].

2. Experience in design and implementation of analytical tools

2.1 Tools for identifying transcriptomic signatures

We will leverage our extensive experience processing and analyzing transcriptomic data to address the aims of the DIRC DIAC. In previous work, we evaluated several independent computational methods and protocols for exon identification, transcript reconstruction and expression level quantification from RNA-Seq data [34]. Our results characterize the strengths and weaknesses of these methods, which would aid the design of analytical strategies.

Following transcriptomic data processing, we can conduct several downstream analyses to identify the functional and regulatory implications of the observed gene expression patterns. 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” [35]. 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, Loregic, computationally integrates gene expression and regulatory network data to characterize the cooperativity of regulatory factors [36]. Loregic uses all 16 possible two-input-one-output logic gates (e.g., AND or XOR) to describe triplets of two factors regulating a common target. The tool finds the gate that best matches each triplet’s observed gene expression pattern across many conditions. Loregic is able to characterize complex circuits involving both transcription factors (TFs) and miRNAs. Additionally, we can exploit cross-species data by using OrthoClust, a computational framework for simultaneously clustering data across multiple species [37]. This tool integrates the co-association networks of individual species by utilizing the orthology relationships of genes between species, and then outputs optimized cross-species modules, either conserved or species specific. A potential application of cross-species modules is to infer putative analogous functions of uncharacterized elements like ncRNAs based on guilt-by-association.

2.2 Tools for the deconvolution of bulk data

Deconvolution refers to the decomposition of a dataset into its constituent components. In exRNA studies, deconvolution methods can help us identify fractions in the bulk expression data and their characteristic expression patterns. We have previously employed several deconvolution analysis methods that can be integrated into the exRNA pipeline, in order to specify subtypes of cells associated with signatures of interest.

We have employed two approaches to the bulk tissue deconvolution (Submitted): an unsupervised approach called non-negative matrix factorization (NMF) and a supervised approach called cell-signature-based decomposition. Given the number of desired components, the bulk tissue gene expression matrix X is decomposed into the product of two matrices: H represents NMF "top components" (NMF-TCs) and V represents the expression level of genes in the NMF-TCs. We found that NMF-TCs recovered the expression patterns of different cell types in bulk RNA-Seq data on brain cell population. We then applied a supervised approach that uses single-cell expression signatures to find the fractions of different cell types. We defined the sample gene expression matrix B, and fraction gene expression matrix iC. We used the non-negative least square method to find a non-negative matrix W as the linear combination coefficients. By applying this method to bulk RNA-Seq data on a brain cell population, we identified cell-fraction changes associated with different traits.

2.3 Tools for network analysis & visualization

We have demonstrated experience in biological network science. After we identify the functional and regulatory networks using the aforementioned pipelines and tools, we will quantify and visualize the properties of these networks to identify possible signatures of dysregulation in the transition from acute to chronic pain.

Our lab has developed various tools for network analysis from multiple perspectives. We and others have used these tools to analyze the human regulatory network [38], the phosphorylation network in yeast [39], the yeast regulatory network [40] and other model organism networks [41]. We have performed and published extensive comparisons between these regulatory networks [42].

TopNet is an automated web tool designed to calculate topological parameters and compare different sub-networks for any given network [43]. This tool computes a variety of topological parameters given the input network and specified subnetworks and calculates the power-law degree distribution for each sub-network. In addition, we developed the TopNet-like Yale Network Analyzer, a web system for managing, comparing and mining multiple networks, which efficiently implements methods that are useful in network analysis [44].

We have also published several papers on constructing hierarchy structures for the regulatory network for both transcriptional and post-transcriptional regulation. We proposed the hierarchical score maximization algorithm, which first defines a score to quantify the degree of hierarchy in a network, and then performs a simulated annealing procedure to infer a hierarchical structure that maximizes the score [41]. We applied our algorithm to determine the hierarchical structure of the phosphorylome in detail. Using genome-wide binding locations of human, worm and fly transcription-regulatory factors (RFs), we performed simulated annealing to reveal the organization of RFs in three layers of master regulators, intermediate regulators and low-level regulators [45]. We organized the binding profiles of 119 TFs in 458 chromatin immunoprecipitation sequencing (ChIP-Seq) experiments from ENCODE into a hierarchy and integrated it with other genomic information (e.g., miRNA regulation), forming a dense meta-network [38].

3. Experience in consortium analyses

### 3..1 Experience in consortium analyses for general genomics

#### **3.1.1 Integrative analysis of consortium-wide datasets**

We played a lead role in the integrative analysis of multi-omic datasets from the [38, 45-48] and modENCODE [48, 49] consortia. By integrating large-scale RNA-Seq and ChIP-Seq datasets from ENCODE, we developed statistical models to quantify the relationship between gene expression and TF binding and/or chromatin modification signatures [50, 51]. We have also developed approaches for constructing and studying biological networks that can be applied to analyze ENCODE datasets. We integrated multiple genomic datasets to construct gene regulatory networks consisting of various regulatory factors including TFs and miRNAs and their target genes [38, 45, 52]. For constructed gene regulatory networks, we developed methods to construct and analyze human and model organism gene regulatory networks [38, 42, 49, 52, 53] using ENCODE and modENCODE datasets. We also analyzed hierarchical structures of gene regulatory networks and found that hierarchy rather than centrality ("hubiness") better reflects the importance of regulators [38, 54-57].

We helped lead the structural variation (SV) analysis for the 1,000 Genomes Project [58-60]. We developed an annotation pipeline that maps single-nucleotide polymorphisms, insertions and deletions (indels) and SVs onto protein-coding genes[61]. We also developed algorithms to identify indels and SVs based on split-read, read-depth and paired-end mapping methods. We studied the distinct features of SVs originating from different mechanisms [62]. We performed SV mechanism annotations for the 1,000 Genomes Project Phase 3 deletions using BreakSeq [62].

We have been an integral part of the Data Integration and Analysis Component (DIAC) for the  Data Management and Resource Repository (DMRR) for the NIH Common Fund Extracellular RNA Communication Consortium [63, 64].

In addition, we participated in the United States Department of Energy Systems Biology Knowledgebase [65], which is an open-source software and data platform that enables data sharing, integration and analysis of microbes, plants and their communities; and the Northeast Structural Genomics Consortium [66], which employs both X-ray crystallography and NMR spectroscopy to provide novel structural information useful in modeling thousands protein domains.

#### **3.1.2 Integrative analysis of omics datasets with genomic variants**

We have extensively analyzed patterns of variation [38, 67, 68]. In recent projects [59, 69], we integrated multiple methods into a comprehensive prioritization pipeline called FunSeq (**Figure 7**). The pipeline identifies sensitive regions with annotations under high selective pressure, links non-coding mutations to their target genes and prioritizes variants based on network connectivity. Recently, we developed RADAR by extending the FunSeq variant prioritization framework to the RNA transcript (In press). RADAR integrates the ENCODE enhanced CLIP datasets, Bind-n-Seq datasets and RBP KD RNA-Seq datasets to reconstruct a comprehensive post-transcriptional network. By combining other genomic information including conservation and motif features, RADAR can pinpoint deleterious variants, such as splicing-disruptive ones, that may be missed by other methods. Finally, we developed a computational tool to systematically annotate upstream open reading frames (uORFs) in the genome [70]. We applied this tool to predict the consequences of genomic variants and somatic mutations on uORFs.

Additionally, we have developed a variety of tools that prioritize protein-coding variants. The Variant Annotation Tool characterizes variants according to affected genes and transcript isoforms [71], and the Analysis of Loss of Function Transcripts tool predicts loss-of-function mutations and their impact [61]. Related, our NetSNP biological network integration tool [72] identifies cancer genes based on connectivity. STRESS [73] and Frustration [74] are two other tools we built to identify mutations that affect allosteric hotspots in proteins and identify key functional protein regions prone to genetic alterations. Our Intensification tool searches for deleterious mutations within repeat regions of proteins [75].

### 3.2 Experience in consortium analyses for disease genomics

#### **3.2.1 Brain diseases**

We played a lead role in the data analysis for the PsychENCODE Consortium [76, 77], a project aimed at understanding regulatory variants in the context of their functional connections to psychiatric disorders, with several papers currently in the revision stage (Submitted). In our recent work, we identified functional elements, multiple QTLs and regulatory-network linkages specific to the adult brain by integrating data from the PsychENCODE Consortium together with relevant external data sources from ENCODE, CommonMind, GTEx and Roadmap (Submitted). In addition to the adult brain, we assessed the degree of chromatin differences between developmental stages relative to that between tissues. Furthermore, we used the regulatory network based on Hi-C, QTLs and activity relationships to connect non-coding genome-wide association study loci to potential psychiatric disease genes including schizophrenia, autism, bipolar disorder and Alzheimer's disease. We also participate in the BrainSpan Consortium, which aims to create a comprehensive map of gene expression and to understand how the human brain changes throughout life. We analyzed large amounts of RNA-Seq data to characterize the transcriptome of the human brain during development [78].

Particularly, we developed an integrated and interpretable deep-learning model, the Deep Structured Phenotype Network (DSPN), that can predict psychiatric disorders using genotype and functional genomic data (Submitted). The model used a Deep Boltzmann Machine (DBM) architecture [79] and included layers for intermediate molecular phenotypes (expression and chromatin state) and pre-defined gene groupings (cell-type markers and co-expression modules), multiple higher layers for inferred groupings (hidden nodes), and a top layer for observed traits (psychiatric disorders and other phenotypes). Finally, we used sparse inter- and intra-level connectivity to integrate QTLs, regulatory networks and co-expression modules.

#### **3.2.2 Cancer**

In addition to neurogenomics and psychiatric diseases, we also have focused on cancer through our role in analyzing data for the Pan-Cancer Analysis Working Group (PCAWG) Consortium [80, 81] and our participation in the cancer genome atlas (TCGA) prostate adenocarcinoma and kidney chromophobe projects [82-84]. We are co-leaders of the PCAWG-2 group, and participate in the analyses of the PCAWG-3, 8 and 11 groups. We leveraged our expertise in non-coding regions in the first whole-genome analysis of TCGA kidney renal papillary cell carcinoma (KIRP) samples, in which we found significant genomic non-coding alterations beyond traditional known drivers of KIRP located within coding exons [82].

We also developed a variety of tools for integrative analysis of cancer genomics data. We developed LARVA, a statistical method for identifying significant mutation enrichments in non-coding elements [85]. Furthermore, we developed MOAT, an alternative empirical mutation burden approach that evaluates mutation enrichments based upon permutations of the input data [86].

1. Habegger L, Sboner A, Gianoulis TA, Rozowsky J, Agarwal A, Snyder M, et al. RSEQtools: a modular framework to analyze RNA-Seq data using compact, anonymized data summaries. Bioinformatics. 2011;27(2):281-3. Epub 2010/12/08. doi: 10.1093/bioinformatics/btq643. PubMed PMID: 21134889; PubMed Central PMCID: PMCPMC3018817.

2. Jee J, Rozowsky J, Yip KY, Lochovsky L, Bjornson R, Zhong G, et al. ACT: aggregation and correlation toolbox for analyses of genome tracks. Bioinformatics. 2011;27(8):1152-4. Epub 2011/02/26. doi: 10.1093/bioinformatics/btr092. PubMed PMID: 21349863; PubMed Central PMCID: PMCPMC3072554.

3. Lu ZJ, Yip KY, Wang G, Shou C, Hillier LW, Khurana E, et al. Prediction and characterization of noncoding RNAs in C. elegans by integrating conservation, secondary structure, and high-throughput sequencing and array data. Genome Res. 2011;21(2):276-85. Epub 2010/12/24. doi: 10.1101/gr.110189.110. PubMed PMID: 21177971; PubMed Central PMCID: PMCPMC3032931.

4. Sboner A, Habegger L, Pflueger D, Terry S, Chen DZ, Rozowsky JS, et al. FusionSeq: a modular framework for finding gene fusions by analyzing paired-end RNA-sequencing data. Genome Biol. 2010;11(10):R104. Epub 2010/10/23. doi: 10.1186/gb-2010-11-10-r104. PubMed PMID: 20964841; PubMed Central PMCID: PMCPMC3218660.

5. Du J, Leng J, Habegger L, Sboner A, McDermott D, Gerstein M. IQSeq: integrated isoform quantification analysis based on next-generation sequencing. PLoS One. 2012;7(1):e29175. Epub 2012/01/13. doi: 10.1371/journal.pone.0029175. PubMed PMID: 22238592; PubMed Central PMCID: PMCPMC3253133.

6. Rozowsky J, Abyzov A, Wang J, Alves P, Raha D, Harmanci A, et al. AlleleSeq: analysis of allele-specific expression and binding in a network framework. Mol Syst Biol. 2011;7:522. Epub 2011/08/04. doi: 10.1038/msb.2011.54. PubMed PMID: 21811232; PubMed Central PMCID: PMCPMC3208341.

7. Pei B, Sisu C, Frankish A, Howald C, Habegger L, Mu XJ, et al. The GENCODE pseudogene resource. Genome Biol. 2012;13(9):R51. Epub 2012/09/07. doi: 10.1186/gb-2012-13-9-r51. PubMed PMID: 22951037; PubMed Central PMCID: PMCPMC3491395.

8. Sisu C, Pei B, Leng J, Frankish A, Zhang Y, Balasubramanian S, et al. Comparative analysis of pseudogenes across three phyla. Proc Natl Acad Sci U S A. 2014;111(37):13361-6. Epub 2014/08/27. doi: 10.1073/pnas.1407293111. PubMed PMID: 25157146; PubMed Central PMCID: PMCPMC4169933.

9. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357-9. Epub 2012/03/06. doi: 10.1038/nmeth.1923. PubMed PMID: 22388286; PubMed Central PMCID: PMCPMC3322381.

10. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics. 2009;25(9):1105-11. Epub 2009/03/18. doi: 10.1093/bioinformatics/btp120. PubMed PMID: 19289445; PubMed Central PMCID: PMCPMC2672628.

11. Sboner A, Karpikov A, Chen G, Smith M, Mattoon D, Freeman-Cook L, et al. Robust-linear-model normalization to reduce technical variability in functional protein microarrays. J Proteome Res. 2009;8(12):5451-64. Epub 2009/10/13. doi: 10.1021/pr900412k. PubMed PMID: 19817483.

12. Smith A, Cheung K, Krauthammer M, Schultz M, Gerstein M. Leveraging the structure of the Semantic Web to enhance information retrieval for proteomics. Bioinformatics. 2007;23(22):3073-9. Epub 2007/10/10. doi: 10.1093/bioinformatics/btm452. PubMed PMID: 17923450.

13. Vidal M, Chan DW, Gerstein M, Mann M, Omenn GS, Tagle D, et al. The human proteome - a scientific opportunity for transforming diagnostics, therapeutics, and healthcare. Clin Proteomics. 2012;9(1):6. Epub 2012/05/16. doi: 10.1186/1559-0275-9-6. PubMed PMID: 22583803; PubMed Central PMCID: PMCPMC3388576.

14. Kitchen RR, Rozowsky JS, Gerstein MB, Nairn AC. Decoding neuroproteomics: integrating the genome, translatome and functional anatomy. Nat Neurosci. 2014;17(11):1491-9. Epub 2014/10/29. doi: 10.1038/nn.3829. PubMed PMID: 25349915; PubMed Central PMCID: PMCPMC4737617.

15. Wu L, Hwang SI, Rezaul K, Lu LJ, Mayya V, Gerstein M, et al. Global survey of human T leukemic cells by integrating proteomics and transcriptomics profiling. Mol Cell Proteomics. 2007;6(8):1343-53. Epub 2007/05/24. doi: 10.1074/mcp.M700017-MCP200. PubMed PMID: 17519225.

16. Yu EZ, Burba AE, Gerstein M. PARE: a tool for comparing protein abundance and mRNA expression data. BMC Bioinformatics. 2007;8:309. Epub 2007/08/28. doi: 10.1186/1471-2105-8-309. PubMed PMID: 17718915; PubMed Central PMCID: PMCPMC2000916.

17. Carlyle BC, Kitchen RR, Zhang J, Wilson RS, Lam TT, Rozowsky JS, et al. Isoform-Level Interpretation of High-Throughput Proteomics Data Enabled by Deep Integration with RNA-seq. J Proteome Res. 2018;17(10):3431-44. Epub 2018/08/21. doi: 10.1021/acs.jproteome.8b00310. PubMed PMID: 30125121.

18. Lin N, Wu B, Jansen R, Gerstein M, Zhao H. Information assessment on predicting protein-protein interactions. BMC Bioinformatics. 2004;5:154. Epub 2004/10/20. doi: 10.1186/1471-2105-5-154. PubMed PMID: 15491499; PubMed Central PMCID: PMCPMC529436.

19. Jansen R, Yu H, Greenbaum D, Kluger Y, Krogan NJ, Chung S, et al. A Bayesian networks approach for predicting protein-protein interactions from genomic data. Science. 2003;302(5644):449-53. Epub 2003/10/18. doi: 10.1126/science.1087361. PubMed PMID: 14564010.

20. Greenbaum D, Colangelo C, Williams K, Gerstein M. Comparing protein abundance and mRNA expression levels on a genomic scale. Genome Biol. 2003;4(9):117. Epub 2003/09/04. doi: 10.1186/gb-2003-4-9-117. PubMed PMID: 12952525; PubMed Central PMCID: PMCPMC193646.

21. Li S, Sullivan NL, Rouphael N, Yu T, Banton S, Maddur MS, et al. Metabolic Phenotypes of Response to Vaccination in Humans. Cell. 2017;169(5):862-77 e17. Epub 2017/05/16. doi: 10.1016/j.cell.2017.04.026. PubMed PMID: 28502771; PubMed Central PMCID: PMCPMC5711477.

22. Gupta S, Kihara Y, Maurya MR, Norris PC, Dennis EA, Subramaniam S. Computational Modeling of Competitive Metabolism between omega3- and omega6-Polyunsaturated Fatty Acids in Inflammatory Macrophages. J Phys Chem B. 2016;120(33):8346-53. Epub 2016/04/12. doi: 10.1021/acs.jpcb.6b02036. PubMed PMID: 27063350; PubMed Central PMCID: PMCPMC5024554.

23. Narasimhan R, Coras R, Rosenthal SB, Sweeney SR, Lodi A, Tiziani S, et al. Serum metabolomic profiling predicts synovial gene expression in rheumatoid arthritis. Arthritis Res Ther. 2018;20(1):164. Epub 2018/08/05. doi: 10.1186/s13075-018-1655-3. PubMed PMID: 30075744; PubMed Central PMCID: PMCPMC6091066.

24. Quehenberger O, Dennis EA. The human plasma lipidome. N Engl J Med. 2011;365(19):1812-23. Epub 2011/11/11. doi: 10.1056/NEJMra1104901. PubMed PMID: 22070478; PubMed Central PMCID: PMCPMC3412394.

25. Quehenberger O, Armando AM, Brown AH, Milne SB, Myers DS, Merrill AH, et al. Lipidomics reveals a remarkable diversity of lipids in human plasma. J Lipid Res. 2010;51(11):3299-305. Epub 2010/07/31. doi: 10.1194/jlr.M009449. PubMed PMID: 20671299; PubMed Central PMCID: PMCPMC2952570.

26. Dennis EA, Deems RA, Harkewicz R, Quehenberger O, Brown HA, Milne SB, et al. A mouse macrophage lipidome. J Biol Chem. 2010;285(51):39976-85. Epub 2010/10/07. doi: 10.1074/jbc.M110.182915. PubMed PMID: 20923771; PubMed Central PMCID: PMCPMC3000979.

27. Gorden DL, Myers DS, Ivanova PT, Fahy E, Maurya MR, Gupta S, et al. Biomarkers of NAFLD progression: a lipidomics approach to an epidemic. J Lipid Res. 2015;56(3):722-36. Epub 2015/01/20. doi: 10.1194/jlr.P056002. PubMed PMID: 25598080; PubMed Central PMCID: PMCPMC4340319.

28. Dumlao DS, Cunningham AM, Wax LE, Norris PC, Hanks JH, Halpin R, et al. Dietary fish oil substitution alters the eicosanoid profile in ankle joints of mice during Lyme infection. J Nutr. 2012;142(8):1582-9. Epub 2012/06/15. doi: 10.3945/jn.112.157883. PubMed PMID: 22695969; PubMed Central PMCID: PMCPMC3397342.

29. Tam VC, Quehenberger O, Oshansky CM, Suen R, Armando AM, Treuting PM, et al. Lipidomic profiling of influenza infection identifies mediators that induce and resolve inflammation. Cell. 2013;154(1):213-27. Epub 2013/07/06. doi: 10.1016/j.cell.2013.05.052. PubMed PMID: 23827684; PubMed Central PMCID: PMCPMC3753192.

30. Dumlao DS, Buczynski MW, Norris PC, Harkewicz R, Dennis EA. High-throughput lipidomic analysis of fatty acid derived eicosanoids and N-acylethanolamines. Biochim Biophys Acta. 2011;1811(11):724-36. Epub 2011/06/22. doi: 10.1016/j.bbalip.2011.06.005. PubMed PMID: 21689782; PubMed Central PMCID: PMCPMC3205334.

31. Wang Y, Armando AM, Quehenberger O, Yan C, Dennis EA. Comprehensive ultra-performance liquid chromatographic separation and mass spectrometric analysis of eicosanoid metabolites in human samples. J Chromatogr A. 2014;1359:60-9. Epub 2014/07/31. doi: 10.1016/j.chroma.2014.07.006. PubMed PMID: 25074422; PubMed Central PMCID: PMCPMC4592635.

32. Dennis EA, Norris PC. Eicosanoid storm in infection and inflammation. Nat Rev Immunol. 2015;15(8):511-23. Epub 2015/07/04. doi: 10.1038/nri3859. PubMed PMID: 26139350; PubMed Central PMCID: PMCPMC4606863.

33. Quehenberger O, Dahlberg-Wright S, Jiang J, Armando AM, Dennis EA. Quantitative determination of esterified eicosanoids and related oxygenated metabolites after base hydrolysis. J Lipid Res. 2018. Epub 2018/10/17. doi: 10.1194/jlr.D089516. PubMed PMID: 30323111.

34. Steijger T, Abril JF, Engstrom PG, Kokocinski F, Consortium R, Hubbard TJ, et al. Assessment of transcript reconstruction methods for RNA-seq. Nat Methods. 2013;10(12):1177-84. Epub 2013/11/05. doi: 10.1038/nmeth.2714. PubMed PMID: 24185837; PubMed Central PMCID: PMCPMC3851240.

35. 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; PubMed Central PMCID: PMCPMC5070849.

36. Wang D, Yan KK, Sisu C, Cheng C, Rozowsky J, Meyerson W, et al. Loregic: a method to characterize the cooperative logic of regulatory factors. PLoS Comput Biol. 2015;11(4):e1004132. Epub 2015/04/18. doi: 10.1371/journal.pcbi.1004132. PubMed PMID: 25884877; PubMed Central PMCID: PMCPMC4401777.

37. Yan KK, Wang D, Rozowsky J, Zheng H, Cheng C, Gerstein M. OrthoClust: an orthology-based network framework for clustering data across multiple species. Genome Biol. 2014;15(8):R100. Epub 2014/09/25. doi: 10.1186/gb-2014-15-8-r100. PubMed PMID: 25249401; PubMed Central PMCID: PMCPMC4289247.

38. Gerstein MB, Kundaje A, Hariharan M, Landt SG, Yan KK, Cheng C, et al. 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; PubMed Central PMCID: PMCPMC4154057.

39. Ptacek J, Devgan G, Michaud G, Zhu H, Zhu X, Fasolo J, et al. Global analysis of protein phosphorylation in yeast. Nature. 2005;438(7068):679-84. Epub 2005/12/02. doi: 10.1038/nature04187. PubMed PMID: 16319894.

40. Krogan NJ, Cagney G, Yu H, Zhong G, Guo X, Ignatchenko A, et al. Global landscape of protein complexes in the yeast Saccharomyces cerevisiae. Nature. 2006;440(7084):637-43. Epub 2006/03/24. doi: 10.1038/nature04670. PubMed PMID: 16554755.

41. Cheng C, Andrews E, Yan KK, Ung M, Wang D, Gerstein M. An approach for determining and measuring network hierarchy applied to comparing the phosphorylome and the regulome. Genome Biol. 2015;16:63. Epub 2015/04/17. doi: 10.1186/s13059-015-0624-2. PubMed PMID: 25880651; PubMed Central PMCID: PMCPMC4404648.

42. Yan KK, Fang G, Bhardwaj N, Alexander RP, Gerstein M. 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. 2010;107(20):9186-91. Epub 2010/05/05. doi: 10.1073/pnas.0914771107. PubMed PMID: 20439753; PubMed Central PMCID: PMCPMC2889091.

43. Yu H, Zhu X, Greenbaum D, Karro J, Gerstein M. TopNet: a tool for comparing biological sub-networks, correlating protein properties with topological statistics. Nucleic Acids Res. 2004;32(1):328-37. Epub 2004/01/16. doi: 10.1093/nar/gkh164. PubMed PMID: 14724320; PubMed Central PMCID: PMCPMC373274.

44. Yip KY, Yu H, Kim PM, Schultz M, Gerstein M. The tYNA platform for comparative interactomics: a web tool for managing, comparing and mining multiple networks. Bioinformatics. 2006;22(23):2968-70. Epub 2006/10/06. doi: 10.1093/bioinformatics/btl488. PubMed PMID: 17021160.

45. Boyle AP, Araya CL, Brdlik C, Cayting P, Cheng C, Cheng Y, et al. 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; PubMed Central PMCID: PMCPMC4336544.

46. Consortium EP. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489(7414):57-74. Epub 2012/09/08. doi: 10.1038/nature11247. PubMed PMID: 22955616; PubMed Central PMCID: PMCPMC3439153.

47. Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, et al. Landscape of transcription in human cells. Nature. 2012;489(7414):101-8. Epub 2012/09/08. doi: 10.1038/nature11233. PubMed PMID: 22955620; PubMed Central PMCID: PMCPMC3684276.

48. Gerstein MB, Rozowsky J, Yan KK, Wang D, Cheng C, Brown JB, et al. 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; PubMed Central PMCID: PMCPMC4155737.

49. Gerstein MB, Lu ZJ, Van Nostrand EL, Cheng C, Arshinoff BI, Liu T, et al. 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; PubMed Central PMCID: PMCPMC3142569.

50. Cheng C, Gerstein M. Modeling the relative relationship of transcription factor binding and histone modifications to gene expression levels in mouse embryonic stem cells. Nucleic Acids Res. 2012;40(2):553-68. Epub 2011/09/20. doi: 10.1093/nar/gkr752. PubMed PMID: 21926158; PubMed Central PMCID: PMCPMC3258143.

51. Cheng C, Alexander R, Min R, Leng J, Yip KY, Rozowsky J, et al. Understanding transcriptional regulation by integrative analysis of transcription factor binding data. Genome Res. 2012;22(9):1658-67. Epub 2012/09/08. doi: 10.1101/gr.136838.111. PubMed PMID: 22955978; PubMed Central PMCID: PMCPMC3431483.

52. Cheng C, Yan KK, Hwang W, Qian J, Bhardwaj N, Rozowsky J, et al. Construction and analysis of an integrated regulatory network derived from high-throughput sequencing data. PLoS Comput Biol. 2011;7(11):e1002190. Epub 2011/11/30. doi: 10.1371/journal.pcbi.1002190. PubMed PMID: 22125477; PubMed Central PMCID: PMCPMC3219617.

53. Negre N, Brown CD, Ma L, Bristow CA, Miller SW, Wagner U, et al. A cis-regulatory map of the Drosophila genome. Nature. 2011;471(7339):527-31. Epub 2011/03/25. doi: 10.1038/nature09990. PubMed PMID: 21430782; PubMed Central PMCID: PMCPMC3179250.

54. Bhardwaj N, Carson MB, Abyzov A, Yan KK, Lu H, Gerstein MB. Analysis of combinatorial regulation: scaling of partnerships between regulators with the number of governed targets. PLoS Comput Biol. 2010;6(5):e1000755. Epub 2010/06/05. doi: 10.1371/journal.pcbi.1000755. PubMed PMID: 20523742; PubMed Central PMCID: PMCPMC2877725.

55. Bhardwaj N, Yan KK, Gerstein MB. 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. 2010;107(15):6841-6. Epub 2010/03/31. doi: 10.1073/pnas.0910867107. PubMed PMID: 20351254; PubMed Central PMCID: PMCPMC2872381.

56. Bhardwaj N, Gerstein M, Lu H. Genome-wide sequence-based prediction of peripheral proteins using a novel semi-supervised learning technique. BMC Bioinformatics. 2010;11 Suppl 1:S6. Epub 2010/03/05. doi: 10.1186/1471-2105-11-S1-S6. PubMed PMID: 20122235; PubMed Central PMCID: PMCPMC3009533.

57. Yu H, Gerstein M. Genomic analysis of the hierarchical structure of regulatory networks. Proc Natl Acad Sci U S A. 2006;103(40):14724-31. Epub 2006/09/28. doi: 10.1073/pnas.0508637103. PubMed PMID: 17003135; PubMed Central PMCID: PMCPMC1595419.

58. Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. Nature. 2015;526(7571):68-74. Epub 2015/10/04. doi: 10.1038/nature15393. PubMed PMID: 26432245; PubMed Central PMCID: PMCPMC4750478.

59. Khurana E, Fu Y, Colonna V, Mu XJ, Kang HM, Lappalainen T, et al. 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; PubMed Central PMCID: PMCPMC3947637.

60. Genomes Project C, Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, et al. 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; PubMed Central PMCID: PMCPMC3042601.

61. Balasubramanian S, Fu Y, Pawashe M, McGillivray P, Jin M, Liu J, et al. Using ALoFT to determine the impact of putative loss-of-function variants in protein-coding genes. Nat Commun. 2017;8(1):382. Epub 2017/08/31. doi: 10.1038/s41467-017-00443-5. PubMed PMID: 28851873; PubMed Central PMCID: PMCPMC5575292.

62. Zhang Y, Li S, Abyzov A, Gerstein MB. Landscape and variation of novel retroduplications in 26 human populations. PLoS Comput Biol. 2017;13(6):e1005567. Epub 2017/07/01. doi: 10.1371/journal.pcbi.1005567. PubMed PMID: 28662076; PubMed Central PMCID: PMCPMC5510864.

63. Freedman JE, Gerstein M, Mick E, Rozowsky J, Levy D, Kitchen R, et al. Diverse human extracellular RNAs are widely detected in human plasma. Nat Commun. 2016;7:11106. Epub 2016/04/27. doi: 10.1038/ncomms11106. PubMed PMID: 27112789; PubMed Central PMCID: PMCPMC4853467.

64. Cheung KH, Keerthikumar S, Roncaglia P, Subramanian SL, Roth ME, Samuel M, et al. Extending gene ontology in the context of extracellular RNA and vesicle communication. J Biomed Semantics. 2016;7:19. Epub 2016/04/15. doi: 10.1186/s13326-016-0061-5. PubMed PMID: 27076901; PubMed Central PMCID: PMCPMC4830068.

65. Arkin AP, Cottingham RW, Henry CS, Harris NL, Stevens RL, Maslov S, et al. KBase: The United States Department of Energy Systems Biology Knowledgebase. Nat Biotechnol. 2018;36(7):566-9. Epub 2018/07/07. doi: 10.1038/nbt.4163. PubMed PMID: 29979655.

66. Huang YJ, Hang D, Lu LJ, Tong L, Gerstein MB, Montelione GT. Targeting the human cancer pathway protein interaction network by structural genomics. Mol Cell Proteomics. 2008;7(10):2048-60. Epub 2008/05/20. doi: 10.1074/mcp.M700550-MCP200. PubMed PMID: 18487680; PubMed Central PMCID: PMCPMC2559933.

67. Yip KY, Cheng C, Bhardwaj N, Brown JB, Leng J, Kundaje A, et al. 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; PubMed Central PMCID: PMCPMC3491392.

68. 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; PubMed Central PMCID: PMCPMC3167619.

69. Fu Y, Liu Z, Lou S, Bedford J, Mu XJ, Yip KY, et al. 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; PubMed Central PMCID: PMCPMC4203974.

70. McGillivray P, Ault R, Pawashe M, Kitchen R, Balasubramanian S, Gerstein M. A comprehensive catalog of predicted functional upstream open reading frames in humans. Nucleic Acids Res. 2018;46(7):3326-38. Epub 2018/03/22. doi: 10.1093/nar/gky188. PubMed PMID: 29562350.

71. Habegger L, Balasubramanian S, Chen DZ, Khurana E, Sboner A, Harmanci A, et al. VAT: a computational framework to functionally annotate variants in personal genomes within a cloud-computing environment. Bioinformatics. 2012;28(17):2267-9. Epub 2012/06/30. doi: 10.1093/bioinformatics/bts368. PubMed PMID: 22743228; PubMed Central PMCID: PMCPMC3426844.

72. Khurana E, Fu Y, Chen J, Gerstein M. Interpretation of genomic variants using a unified biological network approach. PLoS Comput Biol. 2013;9(3):e1002886. Epub 2013/03/19. doi: 10.1371/journal.pcbi.1002886. PubMed PMID: 23505346; PubMed Central PMCID: PMCPMC3591262.

73. Clarke D, Sethi A, Li S, Kumar S, Chang RWF, Chen J, et al. 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; PubMed Central PMCID: PMCPMC4883016.

74. 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; PubMed Central PMCID: PMCPMC5137452.

75. 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; PubMed Central PMCID: PMCPMC5420328.

76. Gandal MJ, Haney JR, Parikshak NN, Leppa V, Ramaswami G, Hartl C, et al. Shared molecular neuropathology across major psychiatric disorders parallels polygenic overlap. Science. 2018;359(6376):693-7. Epub 2018/02/14. doi: 10.1126/science.aad6469. PubMed PMID: 29439242; PubMed Central PMCID: PMCPMC5898828.

77. Psych EC, Akbarian S, Liu C, Knowles JA, Vaccarino FM, Farnham PJ, et al. The PsychENCODE project. Nat Neurosci. 2015;18(12):1707-12. Epub 2015/11/26. doi: 10.1038/nn.4156. PubMed PMID: 26605881; PubMed Central PMCID: PMCPMC4675669.

78. Miller JA, Ding SL, Sunkin SM, Smith KA, Ng L, Szafer A, et al. Transcriptional landscape of the prenatal human brain. Nature. 2014;508(7495):199-206. Epub 2014/04/04. doi: 10.1038/nature13185. PubMed PMID: 24695229; PubMed Central PMCID: PMCPMC4105188.

79. Salakhutdinov R, Hinton G. Deep Boltzmann Machines. In: David van D, Max W, editors. Proceedings of the Twelth International Conference on Artificial Intelligence and Statistics; Proceedings of Machine Learning Research: PMLR; 2009. p. 448--55.

80. Campbell PJ, Getz G, Stuart JM, Korbel JO, Stein LD. Pan-cancer analysis of whole genomes. BioRxiv. 2017:162784.

81. Rodriguez-Martin B, Alvarez EG, Baez-Ortega A, Zamora J, Supek F, Demeulemeester J, et al. Pan-cancer analysis of whole genomes reveals driver rearrangements promoted by LINE-1 retrotransposition in human tumours. BioRxiv. 2018:179705.

82. Li S, Shuch BM, Gerstein MB. Whole-genome analysis of papillary kidney cancer finds significant noncoding alterations. PLoS Genet. 2017;13(3):e1006685. Epub 2017/03/31. doi: 10.1371/journal.pgen.1006685. PubMed PMID: 28358873; PubMed Central PMCID: PMCPMC5391127.

83. Cancer Genome Atlas Research N, Linehan WM, Spellman PT, Ricketts CJ, Creighton CJ, Fei SS, et al. 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; PubMed Central PMCID: PMCPMC4775252.

84. Berger MF, Lawrence MS, Demichelis F, Drier Y, Cibulskis K, Sivachenko AY, et al. The genomic complexity of primary human prostate cancer. Nature. 2011;470(7333):214-20. Epub 2011/02/11. doi: 10.1038/nature09744. PubMed PMID: 21307934; PubMed Central PMCID: PMCPMC3075885.

85. 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; PubMed Central PMCID: PMCPMC4787796.

86. 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; PubMed Central PMCID: PMCPMC5860157.