

# William S. Noble

## List of Publications by Year in descending order

Source: <https://exaly.com/author-pdf/1092884/publications.pdf>

Version: 2024-02-01

229  
papers

51,081  
citations

6613

79  
h-index

1934

207  
g-index

295  
all docs

295  
docs citations

295  
times ranked

60638  
citing authors

| #  | ARTICLE  | IF   | CITATIONS |
|----|--|------|-----------|
| 1  | MEME SUITE: tools for motif discovery and searching. <i>Nucleic Acids Research</i> , 2009, 37, W202-W208.  | 14.5 | 8,180     |
| 2  | Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. <i>Nature</i> , 2007, 447, 799-816.                | 27.8 | 4,709     |
| 3  | FIMO: scanning for occurrences of a given motif. <i>Bioinformatics</i> , 2011, 27, 1017-1018.  | 4.1  | 3,382     |
| 4  | What is a support vector machine?. <i>Nature Biotechnology</i> , 2006, 24, 1565-1567.  | 17.5 | 2,669     |
| 5  | The MEME Suite. <i>Nucleic Acids Research</i> , 2015, 43, W39-W49.   | 14.5 | 2,537     |
| 6  | Semi-supervised learning for peptide identification from shotgun proteomics datasets. <i>Nature Methods</i> , 2007, 4, 923-925.                              | 19.0 | 2,010     |
| 7  | Quantifying similarity between motifs. <i>Genome Biology</i> , 2007, 8, R24.   | 9.6  | 1,579     |
| 8  | Machine learning applications in genetics and genomics. <i>Nature Reviews Genetics</i> , 2015, 16, 321-332.  | 16.3 | 1,334     |
| 9  | Expanded encyclopaedias of DNA elements in the human and mouse genomes. <i>Nature</i> , 2020, 583, 699-710.  | 27.8 | 1,252     |
| 10 | Assessing computational tools for the discovery of transcription factor binding sites. <i>Nature Biotechnology</i> , 2005, 23, 137-144.                      | 17.5 | 1,121     |
| 11 | A three-dimensional model of the yeast genome. <i>Nature</i> , 2010, 465, 363-367.   | 27.8 | 894       |
| 12 | Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. <i>Genome Research</i> , 2012, 22, 1798-1812. | 5.5  | 762       |
| 13 | How does multiple testing correction work?. <i>Nature Biotechnology</i> , 2009, 27, 1135-1137.   | 17.5 | 595       |
| 14 | A statistical framework for genomic data fusion. <i>Bioinformatics</i> , 2004, 20, 2626-2635.  | 4.1  | 568       |
| 15 | Unsupervised pattern discovery in human chromatin structure through genomic segmentation. <i>Nature Methods</i> , 2012, 9, 473-476.                          | 19.0 | 562       |
| 16 | Assigning Significance to Peptides Identified by Tandem Mass Spectrometry Using Decoy Databases. <i>Journal of Proteome Research</i> , 2008, 7, 29-34.       | 3.7  | 551       |
| 17 | Global mapping of protein-DNA interactions in vivo by digital genomic footprinting. <i>Nature Methods</i> , 2009, 6, 283-289.                                | 19.0 | 533       |
| 18 | Integrative annotation of chromatin elements from ENCODE data. <i>Nucleic Acids Research</i> , 2013, 41, 827-841.  | 14.5 | 490       |

| #  | ARTICLE   | IF   | CITATIONS |
|----|---|------|-----------|
| 19 | Kernel methods for predicting protein-protein interactions. <i>Bioinformatics</i> , 2005, 21, i38-i46.  | 4.1  | 447       |
| 20 | Statistical confidence estimation for Hi-C data reveals regulatory chromatin contacts. <i>Genome Research</i> , 2014, 24, 999-1011.   | 5.5  | 444       |
| 21 | Massively multiplex single-cell Hi-C. <i>Nature Methods</i> , 2017, 14, 263-266.  | 19.0 | 441       |
| 22 | A Genome-wide Framework for Mapping Gene Regulation via Cellular Genetic Screens. <i>Cell</i> , 2019, 176, 377-390.e19.   | 28.9 | 379       |
| 23 | HiCRep: assessing the reproducibility of Hi-C data using a stratum-adjusted correlation coefficient. <i>Genome Research</i> , 2017, 27, 1939-1949.  | 5.5  | 376       |
| 24 | Comparative analysis of metazoan chromatin organization. <i>Nature</i> , 2014, 512, 449-452.  | 27.8 | 363       |
| 25 | Fast and Accurate Protein False Discovery Rates on Large-Scale Proteomics Data Sets with Percolator 3.0. <i>Journal of the American Society for Mass Spectrometry</i> , 2016, 27, 1719-1727.                    | 2.8  | 311       |
| 26 | Genome-scale mapping of DNase I sensitivity in vivo using tiling DNA microarrays. <i>Nature Methods</i> , 2006, 3, 511-518.   | 19.0 | 306       |
| 27 | Combining Pairwise Sequence Similarity and Support Vector Machines for Detecting Remote Protein Evolutionary and Structural Relationships. <i>Journal of Computational Biology</i> , 2003, 10, 857-868.         | 1.6  | 300       |
| 28 | Nucleosome positioning signals in genomic DNA. <i>Genome Research</i> , 2007, 17, 1170-1177.  | 5.5  | 287       |
| 29 | Integrative detection and analysis of structural variation in cancer genomes. <i>Nature Genetics</i> , 2018, 50, 1388-1398.   | 21.4 | 268       |
| 30 | Posterior Error Probabilities and False Discovery Rates: Two Sides of the Same Coin. <i>Journal of Proteome Research</i> , 2008, 7, 40-44.  | 3.7  | 264       |
| 31 | Large-scale identification of yeast integral membrane protein interactions. <i>Proceedings of the National Academy of Sciences of the United States of America</i> , 2005, 102, 12123-12128.                    | 7.1  | 260       |
| 32 | Improvements to the Percolator Algorithm for Peptide Identification from Shotgun Proteomics Data Sets. <i>Journal of Proteome Research</i> , 2009, 8, 3737-3745.  | 3.7  | 251       |
| 33 | The Forkhead transcription factor Hcm1 regulates chromosome segregation genes and fills the S-phase gap in the transcriptional circuitry of the cell cycle. <i>Genes and Development</i> , 2006, 20, 2266-2278. | 5.9  | 250       |
| 34 | Analysis of Peptide MS/MS Spectra from Large-Scale Proteomics Experiments Using Spectrum Libraries. <i>Analytical Chemistry</i> , 2006, 78, 5678-5684.  | 6.5  | 237       |
| 35 | Transmembrane Topology and Signal Peptide Prediction Using Dynamic Bayesian Networks. <i>PLoS Computational Biology</i> , 2008, 4, e1000213.  | 3.2  | 232       |
| 36 | The lncRNA Firre anchors the inactive X chromosome to the nucleolus by binding CTCF and maintains H3K27me3 methylation. <i>Genome Biology</i> , 2015, 16, 52.   | 8.8  | 229       |

| #  | ARTICLE  | IF   | CITATIONS |
|----|--|------|-----------|
| 37 | Escape from X Inactivation Varies in Mouse Tissues. PLoS Genetics, 2015, 11, e1005079.   | 3.5  | 224       |
| 38 | A statistical approach for inferring the 3D structure of the genome. Bioinformatics, 2014, 30, i26-i33.  | 4.1  | 223       |
| 39 | A critical assessment of Mus musculus gene function prediction using integrated genomic evidence. Genome Biology, 2008, 9, S2.   | 9.6  | 214       |
| 40 | Bipartite structure of the inactive mouse X chromosome. Genome Biology, 2015, 16, 152.   | 8.8  | 211       |
| 41 | Classification and Subtype Prediction of Adult Soft Tissue Sarcoma by Functional Genomics. American Journal of Pathology, 2003, 163, 691-700.  | 3.8  | 207       |
| 42 | A New Algorithm for the Evaluation of Shotgun Peptide Sequencing in Proteomics: A Support Vector Machine Classification of Peptide MS/MS Spectra and SEQUEST Scores. Journal of Proteome Research, 2003, 2, 137-146. | 3.7  | 206       |
| 43 | Learning Gene Functional Classifications from Multiple Data Types. Journal of Computational Biology, 2002, 9, 401-411.   | 1.6  | 201       |
| 44 | Motif-based analysis of large nucleotide data sets using MEME-ChIP. Nature Protocols, 2014, 9, 1428-1450.  | 12.0 | 200       |
| 45 | Sequence and chromatin determinants of cell-type-specific transcription factor binding. Genome Research, 2012, 22, 1723-1734.  | 5.5  | 196       |
| 46 | Three-dimensional modeling of the <i>P. falciparum</i> genome during the erythrocytic cycle reveals a strong connection between genome architecture and gene expression. Genome Research, 2014, 24, 974-988.         | 5.5  | 193       |
| 47 | Choosing negative examples for the prediction of protein-protein interactions. BMC Bioinformatics, 2006, 7, S2.  | 2.6  | 189       |
| 48 | Faster SEQUEST Searching for Peptide Identification from Tandem Mass Spectra. Journal of Proteome Research, 2011, 10, 3871-3879.   | 3.7  | 185       |
| 49 | Rapid and Accurate Peptide Identification from Tandem Mass Spectra. Journal of Proteome Research, 2008, 7, 3022-3027.  | 3.7  | 181       |
| 50 | Classification of Clear-Cell Sarcoma as a Subtype of Melanoma by Genomic Profiling. Journal of Clinical Oncology, 2003, 21, 1775-1781.   | 1.6  | 177       |
| 51 | Fine-scale chromatin interaction maps reveal the cis-regulatory landscape of human lincRNA genes. Nature Methods, 2015, 12, 71-78.   | 19.0 | 177       |
| 52 | A Deeper Look into Comet's Implementation and Features. Journal of the American Society for Mass Spectrometry, 2015, 26, 1865-1874.  | 2.8  | 175       |
| 53 | Technical advances in proteomics: new developments in data-independent acquisition. F1000Research, 2016, 5, 419.   | 1.6  | 172       |
| 54 | PECAN: library-free peptide detection for data-independent acquisition tandem mass spectrometry data. Nature Methods, 2017, 14, 903-908.   | 19.0 | 158       |

| #  | ARTICLE  | IF   | CITATIONS |
|----|--|------|-----------|
| 55 | Topologically associating domains and their long-range contacts are established during early G1 coincident with the establishment of the replication-timing program. <i>Genome Research</i> , 2015, 25, 1104-1113. | 5.5  | 157       |
| 56 | Non-parametric estimation of posterior error probabilities associated with peptides identified by tandem mass spectrometry. <i>Bioinformatics</i> , 2008, 24, i42-i48.   | 4.1  | 147       |
| 57 | Analysis methods for studying the 3D architecture of the genome. <i>Genome Biology</i> , 2015, 16, 183.  | 8.8  | 146       |
| 58 | Peptide-Centric Proteome Analysis: An Alternative Strategy for the Analysis of Tandem Mass Spectrometry Data. <i>Molecular and Cellular Proteomics</i> , 2015, 14, 2301-2307.                                      | 3.8  | 146       |
| 59 | The spectrum kernel: a string kernel for SVM protein classification. <i>Pacific Symposium on Biocomputing Pacific Symposium on Biocomputing</i> , 2002, , 564-75.  | 0.7  | 144       |
| 60 | Learning to predict protein-protein interactions from protein sequences. <i>Bioinformatics</i> , 2003, 19, 1875-1881.  | 4.1  | 140       |
| 61 | Integrative phenomics reveals insight into the structure of phenotypic diversity in budding yeast. <i>Genome Research</i> , 2013, 23, 1496-1504.   | 5.5  | 138       |
| 62 | Massively parallel profiling and predictive modeling of the outcomes of CRISPR/Cas9-mediated double-strand break repair. <i>Nucleic Acids Research</i> , 2019, 47, 7989-8003.                                      | 14.5 | 135       |
| 63 | MoMo: discovery of statistically significant post-translational modification motifs. <i>Bioinformatics</i> , 2019, 35, 2774-2782.  | 4.1  | 131       |
| 64 | Crux: Rapid Open Source Protein Tandem Mass Spectrometry Analysis. <i>Journal of Proteome Research</i> , 2014, 13, 4488-4491.  | 3.7  | 130       |
| 65 | Measuring the reproducibility and quality of Hi-C data. <i>Genome Biology</i> , 2019, 20, 57.  | 8.8  | 125       |
| 66 | Identification of higher-order functional domains in the human ENCODE regions. <i>Genome Research</i> , 2007, 17, 917-927.   | 5.5  | 120       |
| 67 | Searching for statistically significant regulatory modules. <i>Bioinformatics</i> , 2003, 19, ii16-ii25.   | 4.1  | 115       |
| 68 | Navigating the pitfalls of applying machine learning in genomics. <i>Nature Reviews Genetics</i> , 2022, 23, 169-181.  | 16.3 | 114       |
| 69 | Predicting Human Nucleosome Occupancy from Primary Sequence. <i>PLoS Computational Biology</i> , 2008, 4, e1000134.  | 3.2  | 111       |
| 70 | Protein ranking: From local to global structure in the protein similarity network. <i>Proceedings of the National Academy of Sciences of the United States of America</i> , 2004, 101, 6559-6563.                  | 7.1  | 109       |
| 71 | <scp>quality</scp>: non-parametric estimation of <i>q</i>-values and posterior error probabilities. <i>Bioinformatics</i> , 2009, 25, 964-966.   | 4.1  | 107       |
| 72 | Mapping 3D genome architecture through in situ DNase Hi-C. <i>Nature Protocols</i> , 2016, 11, 2104-2121.  | 12.0 | 106       |

| #  | ARTICLE  | IF   | CITATIONS |
|----|--|------|-----------|
| 73 | Comprehensive identification and analysis of human accelerated regulatory DNA. <i>Genome Research</i> , 2015, 25, 1245-1255.   | 5.5  | 105       |
| 74 | Dynamics of genome reorganization during human cardiogenesis reveal an RBM20-dependent splicing factory. <i>Nature Communications</i> , 2019, 10, 1538.                                | 12.8 | 104       |
| 75 | Epigenetic priors for identifying active transcription factor binding sites. <i>Bioinformatics</i> , 2012, 28, 56-62.  | 4.1  | 103       |
| 76 | Mammalian X Upregulation Is Associated with Enhanced Transcription Initiation, RNA Half-Life, and MOF-Mediated H4K16 Acetylation. <i>Developmental Cell</i> , 2013, 25, 55-68.         | 7.0  | 103       |
| 77 | Efficient Marginalization to Compute Protein Posterior Probabilities from Shotgun Mass Spectrometry Data. <i>Journal of Proteome Research</i> , 2010, 9, 5346-5357.                    | 3.7  | 102       |
| 78 | Unsupervised segmentation of continuous genomic data. <i>Bioinformatics</i> , 2007, 23, 1424-1426.   | 4.1  | 101       |
| 79 | Inferring Clonal Composition from Multiple Sections of a Breast Cancer. <i>PLoS Computational Biology</i> , 2014, 10, e1003703.  | 3.2  | 100       |
| 80 | Empirical comparison of web-based antimicrobial peptide prediction tools. <i>Bioinformatics</i> , 2017, 33, 1921-1929.   | 4.1  | 98        |
| 81 | Critical decisions in metaproteomics: achieving high confidence protein annotations in a sea of unknowns. <i>ISME Journal</i> , 2017, 11, 309-314.                                     | 9.8  | 93        |
| 82 | HiC-spector: a matrix library for spectral and reproducibility analysis of Hi-C contact maps. <i>Bioinformatics</i> , 2017, 33, 2199-2201.   | 4.1  | 92        |
| 83 | Improving Tandem Mass Spectrum Identification Using Peptide Retention Time Prediction across Diverse Chromatography Conditions. <i>Analytical Chemistry</i> , 2007, 79, 6111-6118.     | 6.5  | 90        |
| 84 | GenomeDISCO: a concordance score for chromosome conformation capture experiments using random walks on contact map graphs. <i>Bioinformatics</i> , 2018, 34, 2701-2707.                | 4.1  | 90        |
| 85 | Mass spectrometrists should search only for peptides they care about. <i>Nature Methods</i> , 2015, 12, 605-608.   | 19.0 | 86        |
| 86 | Widely distributed noncoding purifying selection in the human genome. <i>Proceedings of the National Academy of Sciences of the United States of America</i> , 2007, 104, 12410-12415. | 7.1  | 84        |
| 87 | Changes in genome organization of parasite-specific gene families during the Plasmodium transmission stages. <i>Nature Communications</i> , 2018, 9, 1910.                             | 12.8 | 82        |
| 88 | High-Throughput Single-Cell Sequencing with Linear Amplification. <i>Molecular Cell</i> , 2019, 76, 676-690.e10.   | 9.7  | 82        |
| 89 | Ranking predicted protein structures with support vector regression. <i>Proteins: Structure, Function and Bioinformatics</i> , 2008, 71, 1175-1182.                                    | 2.6  | 80        |
| 90 | Metabolic Control over mTOR-Dependent Diapause-like State. <i>Developmental Cell</i> , 2020, 52, 236-250.e7.   | 7.0  | 79        |

| #   | ARTICLE  | IF   | CITATIONS |
|-----|--|------|-----------|
| 91  | Large-scale prediction of protein-protein interactions from structures. <i>BMC Bioinformatics</i> , 2010, 11, 144.   | 2.6  | 77        |
| 92  | Avocado: a multi-scale deep tensor factorization method learns a latent representation of the human epigenome. <i>Genome Biology</i> , 2020, 21, 81.   | 8.8  | 77        |
| 93  | Predicting the in vivo signature of human gene regulatory sequences. <i>Bioinformatics</i> , 2005, 21, i338-i343.  | 4.1  | 76        |
| 94  | Riboproteomics of the Hepatitis C Virus Internal Ribosomal Entry Site. <i>Journal of Proteome Research</i> , 2004, 3, 949-957.   | 3.7  | 75        |
| 95  | Use of shotgun proteomics for the identification, confirmation, and correction of <i>C. elegans</i> gene annotations. <i>Genome Research</i> , 2008, 18, 1660-1669.  | 5.5  | 75        |
| 96  | Joint annotation of chromatin state and chromatin conformation reveals relationships among domain types and identifies domains of cell-type-specific expression. <i>Genome Research</i> , 2015, 25, 544-557. | 5.5  | 74        |
| 97  | Systematic reconstruction of cellular trajectories across mouse embryogenesis. <i>Nature Genetics</i> , 2022, 54, 328-341.   | 21.4 | 73        |
| 98  | A High-Throughput Screen for Transcription Activation Domains Reveals Their Sequence Features and Permits Prediction by Deep Learning. <i>Molecular Cell</i> , 2020, 78, 890-902.e6.                         | 9.7  | 70        |
| 99  | Fast Open Modification Spectral Library Searching through Approximate Nearest Neighbor Indexing. <i>Journal of Proteome Research</i> , 2018, 17, 3463-3474.  | 3.7  | 69        |
| 100 | An Alignment-Free Metapeptide Strategy for Metaproteomic Characterization of Microbiome Samples Using Shotgun Metagenomic Sequencing. <i>Journal of Proteome Research</i> , 2016, 15, 2697-2705.             | 3.7  | 67        |
| 101 | Form and function of topologically associating genomic domains in budding yeast. <i>Proceedings of the National Academy of Sciences of the United States of America</i> , 2017, 114, E3061-E3070.            | 7.1  | 67        |
| 102 | Predicting Co-Complexed Protein Pairs from Heterogeneous Data. <i>PLoS Computational Biology</i> , 2008, 4, e1000054.  | 3.2  | 65        |
| 103 | A Quick Guide to Organizing Computational Biology Projects. <i>PLoS Computational Biology</i> , 2009, 5, e1000424.   | 3.2  | 64        |
| 104 | Tandem Mass Spectrum Identification via Cascaded Search. <i>Journal of Proteome Research</i> , 2015, 14, 3027-3038.  | 3.7  | 63        |
| 105 | Orientation-dependent Dxz4 contacts shape the 3D structure of the inactive X chromosome. <i>Nature Communications</i> , 2018, 9, 1445.   | 12.8 | 63        |
| 106 | Accurate identification of centromere locations in yeast genomes using Hi-C. <i>Nucleic Acids Research</i> , 2015, 43, 5331-5339.  | 14.5 | 61        |
| 107 | PREDICTD PaRallel Epigenomics Data Imputation with Cloud-based Tensor Decomposition. <i>Nature Communications</i> , 2018, 9, 1402.   | 12.8 | 60        |
| 108 | A review of statistical methods for protein identification using tandem mass spectrometry. <i>Statistics and Its Interface</i> , 2012, 5, 3-20.  | 0.3  | 60        |

| #   | ARTICLE  | IF  | CITATIONS |
|-----|--|-----|-----------|
| 109 | Unsupervised embedding of single-cell Hi-C data. <i>Bioinformatics</i> , 2018, 34, i96-i104.   | 4.1 | 59        |
| 110 | Capturing cell type-specific chromatin compartment patterns by applying topic modeling to single-cell Hi-C data. <i>PLoS Computational Biology</i> , 2020, 16, e1008173.   | 3.2 | 59        |
| 111 | Computational and Statistical Analysis of Protein Mass Spectrometry Data. <i>PLoS Computational Biology</i> , 2012, 8, e1002296.   | 3.2 | 57        |
| 112 | The dynamic three-dimensional organization of the diploid yeast genome. <i>ELife</i> , 2017, 6, .  | 6.0 | 57        |
| 113 | Statistical Calibration of the SEQUEST XCorr Function. <i>Journal of Proteome Research</i> , 2009, 8, 2106-2113.   | 3.7 | 56        |
| 114 | Computing Exact p-values for a Cross-correlation Shotgun Proteomics Score Function. <i>Molecular and Cellular Proteomics</i> , 2014, 13, 2467-2479.  | 3.8 | 56        |
| 115 | Improved False Discovery Rate Estimation Procedure for Shotgun Proteomics. <i>Journal of Proteome Research</i> , 2015, 14, 3148-3161.  | 3.7 | 55        |
| 116 | Multiple dimensions of epigenetic gene regulation in the malaria parasite <i>Plasmodium falciparum</i> . <i>BioEssays</i> , 2015, 37, 182-194.   | 2.5 | 54        |
| 117 | High Resolution Models of Transcription Factor-DNA Affinities Improve In Vitro and In Vivo Binding Predictions. <i>PLoS Computational Biology</i> , 2010, 6, e1000916.   | 3.2 | 53        |
| 118 | Software tools for visualizing Hi-C data. <i>Genome Biology</i> , 2017, 18, 26.  | 8.8 | 53        |
| 119 | Integrated systems biology analysis of KSHV latent infection reveals viral induction and reliance on peroxisome mediated lipid metabolism. <i>PLoS Pathogens</i> , 2017, 13, e1006256.                           | 4.7 | 53        |
| 120 | Sci-Hi-C: A single-cell Hi-C method for mapping 3D genome organization in large number of single cells. <i>Methods</i> , 2020, 170, 61-68.   | 3.8 | 53        |
| 121 | On Using Samples of Known Protein Content to Assess the Statistical Calibration of Scores Assigned to Peptide-Spectrum Matches in Shotgun Proteomics. <i>Journal of Proteome Research</i> , 2011, 10, 2671-2678. | 3.7 | 52        |
| 122 | Identifying multi-locus chromatin contacts in human cells using tethered multiple 3C. <i>BMC Genomics</i> , 2015, 16, 121.   | 2.8 | 51        |
| 123 | Modeling peptide fragmentation with dynamic Bayesian networks for peptide identification. <i>Bioinformatics</i> , 2008, 24, i348-i356.   | 4.1 | 50        |
| 124 | Determining the calibration of confidence estimation procedures for unique peptides in shotgun proteomics. <i>Journal of Proteomics</i> , 2013, 80, 123-131.   | 2.4 | 49        |
| 125 | A Unified Multitask Architecture for Predicting Local Protein Properties. <i>PLoS ONE</i> , 2012, 7, e32235.   | 2.5 | 48        |
| 126 | Distinct epigenetic features of differentiation-regulated replication origins. <i>Epigenetics and Chromatin</i> , 2016, 9, 18.   | 3.9 | 47        |



| #   | ARTICLE  | IF   | CITATIONS |
|-----|--|------|-----------|
| 127 | Chromatin compartment dynamics in a haploinsufficient model of cardiac laminopathy. <i>Journal of Cell Biology</i> , 2019, 218, 2919-2944.   | 5.2  | 46        |
| 128 | On the assessment of statistical significance of three-dimensional colocalization of sets of genomic elements. <i>Nucleic Acids Research</i> , 2012, 40, 3849-3855.  | 14.5 | 45        |
| 129 | A dynamic Bayesian network for identifying protein-binding footprints from single molecule-based sequencing data. <i>Bioinformatics</i> , 2010, 26, i334-i342.   | 4.1  | 43        |
| 130 | MetaGOmics: A Web-Based Tool for Peptide-Centric Functional and Taxonomic Analysis of Metaproteomics Data. <i>Proteomes</i> , 2018, 6, 2.  | 3.5  | 43        |
| 131 | Detecting Cross-Linked Peptides by Searching against a Database of Cross-Linked Peptide Pairs. <i>Journal of Proteome Research</i> , 2010, 9, 2488-2495.   | 3.7  | 41        |
| 132 | Matrix-Matched Calibration Curves for Assessing Analytical Figures of Merit in Quantitative Proteomics. <i>Journal of Proteome Research</i> , 2020, 19, 1147-1153.   | 3.7  | 41        |
| 133 | SCOT: Single-Cell Multi-Omics Alignment with Optimal Transport. <i>Journal of Computational Biology</i> , 2022, 29, 3-18.  | 1.6  | 40        |
| 134 | Extremely Fast and Accurate Open Modification Spectral Library Searching of High-Resolution Mass Spectra Using Feature Hashing and Graphics Processing Units. <i>Journal of Proteome Research</i> , 2019, 18, 3792-3799. | 3.7  | 39        |
| 135 | Independent and complementary methods for large-scale structural analysis of mammalian chromatin. <i>Genome Research</i> , 2007, 17, 928-939.  | 5.5  | 38        |
| 136 | Genomic Interaction Profiles in Breast Cancer Reveal Altered Chromatin Architecture. <i>PLoS ONE</i> , 2013, 8, e73974.  | 2.5  | 38        |
| 137 | A unified encyclopedia of human functional DNA elements through fully automated annotation of 164 human cell types. <i>Genome Biology</i> , 2019, 20, 180.   | 8.8  | 37        |
| 138 | Multiple Functional Categories of Proteins Identified in an in Vitro Cellular Ubiquitin Affinity Extract Using Shotgun Peptide Sequencing. <i>Journal of Proteome Research</i> , 2003, 2, 394-404.                       | 3.7  | 35        |
| 139 | A cross-validation scheme for machine learning algorithms in shotgun proteomics. <i>BMC Bioinformatics</i> , 2012, 13, S3.   | 2.6  | 34        |
| 140 | Trans- and cis-acting effects of Firre on epigenetic features of the inactive X chromosome. <i>Nature Communications</i> , 2020, 11, 6053.   | 12.8 | 33        |
| 141 | Unsupervised manifold alignment for single-cell multi-omics data. , 2020, 2020, 1-10.  |      | 33        |
| 142 | Automated mapping of large-scale chromatin structure in ENCODE. <i>Bioinformatics</i> , 2008, 24, 1911-1916.   | 4.1  | 31        |
| 143 | A genome-wide 3C-method for characterizing the three-dimensional architectures of genomes. <i>Methods</i> , 2012, 58, 277-288.   | 3.8  | 31        |
| 144 | Maintaining replication origins in the face of genomic change. <i>Genome Research</i> , 2012, 22, 1940-1952.   | 5.5  | 31        |

| #   | ARTICLE   | IF   | CITATIONS |
|-----|---|------|-----------|
| 145 | Segway 2.0: Gaussian mixture models and minibatch training. <i>Bioinformatics</i> , 2018, 34, 669-671.  | 4.1  | 31        |
| 146 | On the Importance of Well-Calibrated Scores for Identifying Shotgun Proteomics Spectra. <i>Journal of Proteome Research</i> , 2015, 14, 1147-1160.  | 3.7  | 30        |
| 147 | Calibration Using a Single-Point External Reference Material Harmonizes Quantitative Mass Spectrometry Proteomics Data between Platforms and Laboratories. <i>Analytical Chemistry</i> , 2018, 90, 13112-13117.                 | 6.5  | 30        |
| 148 | A pitfall for machine learning methods aiming to predict across cell types. <i>Genome Biology</i> , 2020, 21, 282.  | 8.8  | 29        |
| 149 | The spring-loaded genome: Nucleosome redistributions are widespread, transient, and DNA-directed. <i>Genome Research</i> , 2014, 24, 251-259.   | 5.5  | 28        |
| 150 | mokapot: Fast and Flexible Semisupervised Learning for Peptide Detection. <i>Journal of Proteome Research</i> , 2021, 20, 1966-1971.  | 3.7  | 28        |
| 151 | Identifying remote protein homologs by network propagation. <i>FEBS Journal</i> , 2005, 272, 5119-5128.   | 4.7  | 27        |
| 152 | A thermodynamic approach to PCR primer design. <i>Nucleic Acids Research</i> , 2009, 37, e95-e95.   | 14.5 | 27        |
| 153 | Computational methods for analyzing and modeling genome structure and organization. <i>Wiley Interdisciplinary Reviews: Systems Biology and Medicine</i> , 2019, 11, e1435.   | 6.6  | 27        |
| 154 | Averaging Strategy To Reduce Variability in Target-Decoy Estimates of False Discovery Rate. <i>Journal of Proteome Research</i> , 2019, 18, 585-593.  | 3.7  | 26        |
| 155 | Protein Family Classification Using Sparse Markov Transducers. <i>Journal of Computational Biology</i> , 2003, 10, 187-213.   | 1.6  | 25        |
| 156 | Param-Medic: A Tool for Improving MS/MS Database Search Yield by Optimizing Parameter Settings. <i>Journal of Proteome Research</i> , 2017, 16, 1817-1824.  | 3.7  | 25        |
| 157 | DNA sequence+shape kernel enables alignment-free modeling of transcription factor binding. <i>Bioinformatics</i> , 2017, 33, 3003-3010.   | 4.1  | 25        |
| 158 | A learned embedding for efficient joint analysis of millions of mass spectra. <i>Nature Methods</i> , 2022, 19, 675-678.  | 19.0 | 25        |
| 159 | Detecting Remote Evolutionary Relationships among Proteins by Large-Scale Semantic Embedding. <i>PLoS Computational Biology</i> , 2011, 7, e1001047.  | 3.2  | 24        |
| 160 | Dealing with large diagonals in kernel matrices. <i>Annals of the Institute of Statistical Mathematics</i> , 2003, 55, 391-408.   | 0.8  | 23        |
| 161 | Combining High-Resolution and Exact Calibration To Boost Statistical Power: A Well-Calibrated Score Function for High-Resolution MS2 Data. <i>Journal of Proteome Research</i> , 2018, 17, 3644-3656.                           | 3.7  | 23        |
| 162 | Predicting gene expression in the human malaria parasite <i>Plasmodium falciparum</i> using histone modification, nucleosome positioning, and 3D localization features. <i>PLoS Computational Biology</i> , 2019, 15, e1007329. | 3.2  | 23        |

| #   | ARTICLE  | IF   | CITATIONS |
|-----|--|------|-----------|
| 163 | Systematic proteomics of endogenous human cohesin reveals an interaction with diverse splicing factors and RNA-binding proteins required for mitotic progression. <i>Journal of Biological Chemistry</i> , 2019, 294, 8760-8772.   | 3.4  | 23        |
| 164 | Completing the ENCODE3 compendium yields accurate imputations across a variety of assays and human biosamples. <i>Genome Biology</i> , 2020, 21, 82.   | 8.8  | 22        |
| 165 | HiCRep.py: fast comparison of Hi-C contact matrices in Python. <i>Bioinformatics</i> , 2021, 37, 2996-2997.  | 4.1  | 22        |
| 166 | Metaproteomics reveal that rapid perturbations in organic matter prioritize functional restructuring over taxonomy in western Arctic Ocean microbiomes. <i>ISME Journal</i> , 2020, 14, 39-52.                                     | 9.8  | 21        |
| 167 | Learning a Weighted Sequence Model of the Nucleosome Core and Linker Yields More Accurate Predictions in <i>Saccharomyces cerevisiae</i> and <i>Homo sapiens</i> . <i>PLoS Computational Biology</i> , 2010, 6, e1000834.          | 3.2  | 20        |
| 168 | Exploratory analysis of genomic segmentations with Segtools. <i>BMC Bioinformatics</i> , 2011, 12, 415.  | 2.6  | 20        |
| 169 | Using DNase Hi-C techniques to map global and local three-dimensional genome architecture at high resolution. <i>Methods</i> , 2018, 142, 59-73.   | 3.8  | 20        |
| 170 | Ten simple rules for writing a response to reviewers. <i>PLoS Computational Biology</i> , 2017, 13, e1005730.  | 3.2  | 20        |
| 171 | Measuring significant changes in chromatin conformation with ACCOST. <i>Nucleic Acids Research</i> , 2020, 48, 2303-2311.  | 14.5 | 19        |
| 172 | Local chromatin fiber folding represses transcription and loop extrusion in quiescent cells. <i>ELife</i> , 2021, 10, .  | 6.0  | 18        |
| 173 | Implications of COMT long-range interactions on the phenotypic variability of 22q11.2 deletion syndrome. <i>Nucleus</i> , 2013, 4, 487-493.  | 2.2  | 17        |
| 174 | Choosing panels of genomics assays using submodular optimization. <i>Genome Biology</i> , 2016, 17, 229.   | 8.8  | 16        |
| 175 | Large-scale tandem mass spectrum clustering using fast nearest neighbor searching. <i>Rapid Communications in Mass Spectrometry</i> , 2021, , e9153.   | 1.5  | 16        |
| 176 | Using machine learning to speed up manual image annotation: application to a 3D imaging protocol for measuring single cell gene expression in the developing <i>C. elegans</i> embryo. <i>BMC Bioinformatics</i> , 2010, 11, 84.   | 2.6  | 15        |
| 177 | The GIPC1-Akt1 Pathway Is Required for the Specification of the Eye Field in Mouse Embryonic Stem Cells. <i>Stem Cells</i> , 2015, 33, 2674-2685.  | 3.2  | 15        |
| 178 | Machine Learning Strategy That Leverages Large Data sets to Boost Statistical Power in Small-Scale Experiments. <i>Journal of Proteome Research</i> , 2020, 19, 1267-1274.   | 3.7  | 15        |
| 179 | Faster Mass Spectrometry-Based Protein Inference: Junction Trees Are More Efficient than Sampling and Marginalization by Enumeration. <i>IEEE/ACM Transactions on Computational Biology and Bioinformatics</i> , 2012, 9, 809-817. | 3.0  | 14        |
| 180 | Conversion of Prostate Adenocarcinoma to Small Cell Carcinoma-Like by Reprogramming. <i>Journal of Cellular Physiology</i> , 2016, 231, 2040-2047.   | 4.1  | 14        |

| #   | ARTICLE   | IF   | CITATIONS |
|-----|---|------|-----------|
| 181 | Analysis of secondary structure in proteins by chemical cross-linking coupled to MS. <i>Proteomics</i> , 2012, 12, 2746-2752.   | 2.2  | 13        |
| 182 | Inducible CRISPR genome editing platform in naive human embryonic stem cells reveals JARID2 function in self-renewal. <i>Cell Cycle</i> , 2018, 17, 00-00.  | 2.6  | 13        |
| 183 | Dynamic Bayesian Network for Accurate Detection of Peptides from Tandem Mass Spectra. <i>Journal of Proteome Research</i> , 2016, 15, 2749-2759.  | 3.7  | 12        |
| 184 | Choosing non-redundant representative subsets of protein sequence data sets using submodular optimization. <i>Proteins: Structure, Function and Bioinformatics</i> , 2018, 86, 454-466.                   | 2.6  | 12        |
| 185 | Multiple Competition-Based FDR Control and Its Application to Peptide Detection. <i>Lecture Notes in Computer Science</i> , 2020, , 54-71.  | 1.3  | 12        |
| 186 | Progressive Calibration and Averaging for Tandem Mass Spectrometry Statistical Confidence Estimation: Why Settle for a Single Decoy?. <i>Lecture Notes in Computer Science</i> , 2017, 10229, 99-116.     | 1.3  | 11        |
| 187 | Single-cell landscape of nuclear configuration and gene expression during stem cell differentiation and X inactivation. <i>Genome Biology</i> , 2021, 22, 279.  | 8.8  | 11        |
| 188 | Identification of genes escaping X inactivation by allelic expression analysis in a novel hybrid mouse model. <i>Data in Brief</i> , 2015, 5, 761-769.  | 1.0  | 10        |
| 189 | Subzero, saline incubations of <i>Colwellia psychrerythraea</i> reveal strategies and biomarkers for sustained life in extreme icy environments. <i>Environmental Microbiology</i> , 2021, 23, 3840-3866. | 3.8  | 10        |
| 190 | Accurately Assigning Peptides to Spectra When Only a Subset of Peptides Are Relevant. <i>Journal of Proteome Research</i> , 2021, 20, 4153-4164.  | 3.7  | 10        |
| 191 | Precursor Charge State Prediction for Electron Transfer Dissociation Tandem Mass Spectra. <i>Journal of Proteome Research</i> , 2010, 9, 5438-5444.   | 3.7  | 9         |
| 192 | Response to "Mass spectrometrists should search for all peptides, but assess only the ones they care about". <i>Nature Methods</i> , 2017, 14, 644-644.   | 19.0 | 9         |
| 193 | DIAMeter: matching peptides to data-independent acquisition mass spectrometry data. <i>Bioinformatics</i> , 2021, 37, i434-i442.  | 4.1  | 9         |
| 194 | ppx: Programmatic Access to Proteomics Data Repositories. <i>Journal of Proteome Research</i> , 2021, 20, 4621-4624.  | 3.7  | 9         |
| 195 | MCAST: scanning for cis-regulatory motif clusters. <i>Bioinformatics</i> , 2016, 32, 1217-1219.   | 4.1  | 8         |
| 196 | Speeding Up Percolator. <i>Journal of Proteome Research</i> , 2019, 18, 3353-3359.  | 3.7  | 8         |
| 197 | Comprehensive characterization of tissue-specific chromatin accessibility in L2 <i>Caenorhabditis elegans</i> nematodes. <i>Genome Research</i> , 2021, 31, 1952-1969.                                    | 5.5  | 8         |
| 198 | Reducing Peptide Sequence Bias in Quantitative Mass Spectrometry Data with Machine Learning. <i>Journal of Proteome Research</i> , 2022, 21, 1771-1782.   | 3.7  | 8         |

| #   | ARTICLE   | IF   | CITATIONS |
|-----|---|------|-----------|
| 199 | Proteomic analyses of nucleus laminaris identified candidate targets of the fragile X mental retardation protein. <i>Journal of Comparative Neurology</i> , 2017, 525, 3341-3359.   | 1.6  | 7         |
| 200 | Controlling the FDR in Imperfect Matches to an Incomplete Database. <i>Journal of the American Statistical Association</i> , 2018, 113, 973-982.  | 3.1  | 7         |
| 201 | Prediction of Phenotype Information from Genotype Data. <i>Communications in Information and Systems</i> , 2010, 10, 99-114.  | 0.5  | 7         |
| 202 | Interpretation of the DOME Recommendations for Machine Learning in Proteomics and Metabolomics. <i>Journal of Proteome Research</i> , 2022, 21, 1204-1207.  | 3.7  | 7         |
| 203 | Meteorological factors and timing of the initiating event of human parturition. <i>International Journal of Biometeorology</i> , 2011, 55, 265-272.   | 3.0  | 6         |
| 204 | Bipartite matching generalizations for peptide identification in tandem mass spectrometry. , 2016, , .  |      | 6         |
| 205 | Detecting Modifications in Proteomics Experiments with Param-Medic. <i>Journal of Proteome Research</i> , 2019, 18, 1902-1906.  | 3.7  | 6         |
| 206 | Building Spectral Libraries from Narrow-Window Data-Independent Acquisition Mass Spectrometry Data. <i>Journal of Proteome Research</i> , 2022, 21, 1382-1391.  | 3.7  | 6         |
| 207 | Visualization and Dissemination of Multidimensional Proteomics Data Comparing Protein Abundance During <i>Caenorhabditis elegans</i> Development. <i>Journal of the American Society for Mass Spectrometry</i> , 2015, 26, 1827-1836. | 2.8  | 5         |
| 208 | Title is missing!. <i>Annals of the Institute of Statistical Mathematics</i> , 2003, 55, 391-408.   | 0.8  | 5         |
| 209 | Semi-supervised Single-Cell Cross-modality Translation Using Polarbear. <i>Lecture Notes in Computer Science</i> , 2022, , 20-35.   | 1.3  | 5         |
| 210 | Joint Precursor Elution Profile Inference via Regression for Peptide Detection in Data-Independent Acquisition Mass Spectra. <i>Journal of Proteome Research</i> , 2019, 18, 86-94.   | 3.7  | 4         |
| 211 | Submodular sketches of single-cell RNA-seq measurements. , 2020, , .  |      | 4         |
| 212 | Learning Score Function Parameters for Improved Spectrum Identification in Tandem Mass Spectrometry Experiments. <i>Journal of Proteome Research</i> , 2012, 11, 4499-4508.   | 3.7  | 3         |
| 213 | Tandem Mass Spectrometry-Based Amyloid Typing Using Manual Microdissection and Open-Source Data Processing. <i>American Journal of Clinical Pathology</i> , 2022, 157, 748-757.   | 0.7  | 3         |
| 214 | Protein Interaction Networks: Protein Domain Interaction and Protein Function Prediction. , 2011, , 427-459.  |      | 3         |
| 215 | Zero-shot imputations across species are enabled through joint modeling of human and mouse epigenomics. , 2020, , .   |      | 3         |
| 216 | A wider field of view to predict expression. <i>Nature Methods</i> , 2021, 18, 1155-1156.   | 19.0 | 3         |

| #   | ARTICLE   | IF  | CITATIONS |
|-----|---|-----|-----------|
| 217 | On the Relationship between DNA Periodicity and Local Chromatin Structure. Lecture Notes in Computer Science, 2009, , 434-450.  | 1.3 | 3         |
| 218 | Predictive model of 3D domain formation via CTCF-mediated extrusion. Proceedings of the National Academy of Sciences of the United States of America, 2015, 112, 14404-14405.   | 7.1 | 2         |
| 219 | Submodular Generalized Matching for Peptide Identification in Tandem Mass Spectrometry. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2019, 16, 1168-1181. | 3.0 | 2         |
| 220 | Prioritizing transcriptomic and epigenomic experiments using an optimization strategy that leverages imputed data. Bioinformatics, 2021, 37, 439-447.                           | 4.1 | 2         |
| 221 | Using Substitution Matrices to Estimate Probability Distributions for Biological Sequences. Journal of Computational Biology, 2002, 9, 775-791.                                 | 1.6 | 1         |
| 222 | Faster and more accurate graphical model identification of tandem mass spectra using trellises. Bioinformatics, 2016, 32, i322-i331.  | 4.1 | 1         |
| 223 | Comprehensive statistical inference of the clonal structure of cancer from multiple biopsies. Scientific Reports, 2017, 7, 16943.   | 3.3 | 1         |
| 224 | Predicting Nucleosome Positioning Using Multiple Evidence Tracks. Lecture Notes in Computer Science, 2010, , 441-455.   | 1.3 | 1         |
| 225 | Avocado. , 2020, , .  |     | 1         |
| 226 | AUTOMATED VALIDATION OF POLYMERASE CHAIN REACTION AMPLICON MELTING CURVES. Journal of Bioinformatics and Computational Biology, 2006, 04, 299-315.                              | 0.8 | 0         |
| 227 | Choosing Non-redundant Representative Subsets Of Protein Sequence Data Sets Using Submodular Optimization. , 2018, , .  |     | 0         |
| 228 | Response to comments on "Empirical comparison of web-based antimicrobial peptide prediction tools". Bioinformatics, 2019, 35, 2695-2696.  | 4.1 | 0         |
| 229 | Submodular Maximization via Gradient Ascent: The Case of Deep Submodular Functions. Advances in Neural Information Processing Systems, 2018, 2018, 7989-7999.                   | 2.8 | 0         |