

PAVEL P. KUKSA

3700 Hamilton Walk, Philadelphia, PA 19104

+1 215-573-3114; pkuksa@pennmedicine.upenn.edu

URL: <https://pkuksa.org/>

Education

- **Ph.D.**—Computer Science, Rutgers University 2011
Dissertation: *Scalable Kernel Methods and Algorithms for General Sequence Analysis*
Advisor: Vladimir Pavlovic
- **M.Sc.**—Computer Science, Rutgers University 2010
Advisor: Vladimir Pavlovic
- **M.Sc. (Honors)**—Computer and Information Sciences, Bauman Moscow State Technical University, Moscow, Russia 2004
- **B.Sc. (Honors)**—Computer Engineering, Bauman Moscow State Technical University, Moscow, Russia 2002

Current Position

- Research Assistant Professor, Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania

Professional Interests

- Biomedical informatics
- Algorithms and methods for high-throughput sequencing and functional genomics data
- Modeling and analysis of sequences and sequencing data
- Large-scale data mining algorithms and inference
- Applied machine learning
- Natural language processing

Professional Experience

- **Research Assistant Professor**—University of Pennsylvania 2022-
- **Research Associate**—University of Pennsylvania 2019-2022
- **Postdoctoral Researcher**—University of Pennsylvania 2014-2019
 - Mentor: Dr. Li-San Wang
- **Postdoctoral Scientist**—NEC Laboratories America, Princeton 2011-2014
 - Machine Learning Department
- **Research Intern**—NEC Laboratories America, Princeton, NJ May 2009 – August 2009
 - Supervisor: Dr. Yanjun Qi
- **Research Intern**—NEC Laboratories America, Princeton, NJ May 2008 – August 2008
 - Supervisor: Dr. Jason Weston
- **Teaching Assistant**—Department of Computer Science, Rutgers University 09/2005-12/2010
- **Software Engineer**—Mobile Business Systems 2003–2005

Publications

Note: In Computer Science, *conferences* and *workshops* (focused expert meetings) are primary venues for publishing original research findings. Top conferences and workshops are highly selective and have *lower acceptance rates* and are *highly regarded*. *Technical reports* officially document results of research projects and are often used to report original findings ahead of formal conference or journal publication.

- **Ph.D. Thesis**

1. **Pavel P. Kuksa**. Scalable Kernel Methods and Algorithms for General Sequence Analysis. Ph.D. thesis, Rutgers University, March 2011. *Dissertation Committee*: Vladimir Pavlovic, Casimir Kulikowski, Alexander Schliep, Christina Leslie DOI: <https://doi.org/doi:10.7282/T3QZ299S>

- **Book chapters**

1. Amlie-Wolf A., **Kuksa Pavel P.**, Lee C.-Y., Mlynarski E., Leung Y.Y., and Wang L.-S. Using INFERNO to infer the molecular mechanisms underlying noncoding genetic associations. In Haiming Cao, editor, *Functional Analysis of Long Non-Coding RNAs*, volume 2254 of *Methods in Molecular Biology*. Humana, New York, NY, 2021. DOI: https://doi.org/10.1007/978-1-0716-1158-6_6
2. **Pavel P. Kuksa**, Yuk Yee Leung, Lee E. Vandivier, Zachary Anderson, Brian D. Gregory, Li-San Wang. *In silico* identification of RNA modifications from high-throughput sequencing data. In *RNA Methylation: Methods and Protocols*, pp. 211-229. Springer, 2017 DOI: https://doi.org/10.1007/978-1-4939-6807-7_14

- **Peer-reviewed Journal Papers**

1. **Pavel P. Kuksa**, Emily Greenfest-Allen, Jeffrey Cifello, Matei Ionita, Hui Wang, Heather Nicaretta, Po-Liang Cheng, Wan-Ping Lee, Li-San Wang, Yuk Yee Leung. Scalable approaches for functional analyses of whole-genome sequencing non-coding variants *Human Molecular Genetics*, Volume 31, Issue R1, 15 October 2022, Pages R62–R72 DOI: <https://doi.org/10.1093/hmg/ddac191>
2. **Pavel P. Kuksa**, Y. Y. Leung, P. Gangadharan, Z. Katanic, L. Kleidermacher, A. Amlie-Wolf, C.-Y. Lee, L. Qu, E. Greenfest-Allen, O. Valladares, and L.-S. Wang. FILER: a framework for harmonizing and querying large-scale functional genomics knowledge. *NAR Genomics and Bioinformatics*, 4(1), Jan 2022. DOI: <https://doi.org/10.1093/nargab/lqab123>
3. **Pavel P. Kuksa**, C.-L. Liu, W. Fu, L. Qu, Y. Zhao, Z. Katanic, K. Clark, A. B. Kuzma, P.-C. Ho, K.-T. Tzeng, O. Valladares, S.-Y. Chou, A. C. Naj, G. D. Schellenberg, L.-S. Wang, and Y. Y. Leung. Alzheimer’s Disease variant portal: A catalog of genetic findings for Alzheimer’s Disease. *Journal of Alzheimer’s Disease*, 86(1):461-477, March 2022. DOI: <https://doi.org/10.3233/JAD-215055>
4. **Pavel P. Kuksa**, Chien-Yueh Lee, Alexandre Amlie-Wolf, Prabhakaran Gangadharan, Elizabeth E Mlynarski, Yi-Fan Chou, Han-Jen Lin, Heather Issen, Emily Greenfest-Allen, Otto Valladares, Yuk Yee Leung, and Li-San Wang. SparkINFERNO: A scalable high-throughput pipeline for inferring molecular mechanisms of non-coding genetic variants. *Bioinformatics*, 2020. DOI: <https://doi.org/10.1093/bioinformatics/btaa246>

5. **Pavel P. Kuksa**, Fan Li, Sampath Kannan, Brian D. Gregory, Yuk Yee Leung, and Li-San Wang. HiPR: High-throughput probabilistic RNA structure inference. *Computational and Structural Biotechnology Journal*, 18:1539–1547, 2020. DOI: <https://doi.org/10.1016/j.csbj.2020.06.004>
6. **Pavel P. Kuksa**, Alexandre Amlie-Wolf, Yih-Chii Hwang, Otto Valladares, Brian D. Gregory, and Li-San Wang. HIPPIE2: a method for fine-scale identification of physically interacting chromatin regions. *NAR Genomics and Bioinformatics*, 2020. DOI: <https://doi.org/10.1093/nargab/lqaa022>
7. **Pavel P. Kuksa**, Alexandre Amlie-Wolf, Zivadin Katanic, Otto Valladares, Li-San Wang, Yuk Yee Leung. DASHR 2.0: integrated database of human small non-coding RNA genes and mature products. *Bioinformatics*, 2019 DOI: <https://doi.org/10.1093/bioinformatics/bty709>
8. **Pavel P. Kuksa**, Alexandre Amlie-Wolf, Zivadin Katanic, Otto Valladares, Li-San Wang, Yuk Yee Leung. SPAR: small RNA-seq portal for analysis of sequencing experiments. *Nucleic Acids Research*, 2018 DOI: <https://doi.org/10.1093/nar/gky330>
9. A. Amlie-Wolf, M. Tang, EE Mlynarski, **Pavel P. Kuksa**, O Valladares, Z Katanic, Debby Tsuang, Christopher D Brown, Gerard D Schellenberg, Li-San Wang. INFERNO: inferring the molecular mechanisms of noncoding genetic variants. *Nucleic Acids Research* 46 (17), pp. 8740-8753, 2018 DOI: <https://doi.org/10.1093/nar/gky686>
10. Yuk Yee Leung*, **Pavel P. Kuksa***, Alexandre Amlie-Wolf, Otto Valladares, Lyle H. Ungar, Sampath Kannan, Brian D. Gregory, and Li-San Wang. DASHR: database of small human noncoding RNAs. *Nucleic Acids Research (Database Issue)*, 2016. DOI: <https://dx.doi.org/10.1093/nar/gkv1188>
11. **Pavel P. Kuksa**, Martin Renqiang Min, Rishabh Dugar, Mark Gerstein. High-order neural networks and kernel methods for peptide-MHC binding prediction. *Bioinformatics*, 2015. DOI: <http://dx.doi.org/10.1093/bioinformatics/btv371>
12. Lee E. Vandivier, Rafael Campos, **Pavel P. Kuksa**, Ian M. Silverman, Li-San Wang, and Brian D. Gregory. Chemical modifications mark alternatively spliced and uncapped messenger RNAs in Arabidopsis. *The Plant Cell*, 27(11):3024–3037, 2015. DOI: <http://dx.doi.org/10.1105/tpc.15.00591>
13. Shawn W. Foley, Lee E. Vandivier, **Pavel P. Kuksa**, Brian D. Gregory. Transcriptome-wide measurement of plant RNA secondary structure. *Current Opinion in Plant Biology*, 2015. DOI: <http://dx.doi.org/10.1016/j.pbi.2015.05.021>
14. Yih-Chii Hwang, Chiao-Feng Lin, Otto Valladares, John Malamon, **Pavel P. Kuksa**, Qi Zheng, Brian D. Gregory, and Li-San Wang. HIPPIE: A high-throughput identification pipeline for promoter interacting enhancer elements. *Bioinformatics*, 2014. DOI: <http://dx.doi.org/10.1093/bioinformatics/btv371>
15. **Pavel P. Kuksa**. Biological sequence analysis with multivariate string kernels. *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)*, March 2013. DOI: <https://doi.org/10.1109/TCBB.2013.15>

16. Ronan Collobert, Jason Weston, Leon Bottou, Michael Karlen, Koray Kavukcuoglu, and **Pavel Kuksa**. Natural language processing (almost) from scratch. *Journal of Machine Learning Research (JMLR)*, 12:2493–2537, 2011. URL: <https://www.jmlr.org/papers/volume12/collobert11a/collobert11a.pdf>
17. **Pavel Kuksa** and Vladimir Pavlovic. Efficient motif finding algorithms for large-alphabet inputs. *BMC Bioinformatics*, 11(Suppl 8):S1, 2010. DOI: <http://dx.doi.org/10.1186/1471-2105-11-S8-S1>
18. **Pavel Kuksa**, Pai-Hsi Huang, and Vladimir Pavlovic. Efficient use of unlabeled data for protein sequence classification: a comparative study. *BMC Bioinformatics*, 10(Suppl 4):S2, 2009. DOI: <http://dx.doi.org/10.1186/1471-2105-10-S4-S2>
19. **Pavel Kuksa** and Vladimir Pavlovic. Efficient alignment-free DNA barcode analytics. *BMC Bioinformatics*, 10(Suppl 14):S9, 2009. DOI: <http://dx.doi.org/10.1186/1471-2105-10-S14-S9>

- **Peer-reviewed Conference Papers**

1. **Pavel P. Kuksa**, Martin Renquiang Min, Rishabh Dugar, and Mark Gerstein. High-Order Neural Networks and Kernel Methods for Peptide-MHC Binding Prediction. In *NIPS Machine Learning in Computational Biology*, 2014.
2. **Pavel P. Kuksa** and Vladimir Pavlovic. Efficient evaluation of large sequence kernels. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 2012. Oral presentation. Acceptance rate: 72/755 (9.5%)
3. **Pavel P. Kuksa**, Imdadullah Khan, and Vladimir Pavlovic. Generalized similarity kernels for efficient sequence classification. In *SIAM International Conference on Data Mining (SDM)*, 2012. Acceptance rate: 99/362 (27%).
4. **Pavel P. Kuksa**. Efficient sequence kernel-based genome-wide prediction of transcription factors. In *International Conference on Pattern Recognition (ICPR)*, 2012.
5. **Pavel P. Kuksa**. 2D similarity kernels for biological sequence classification. In *International Workshop on Data Mining in Bioinformatics (BIOKDD)*, 2012.
6. **Pavel Kuksa** and Yanjun Qi. Semi-supervised bio-named entity recognition with word-codebook learning. In *SIAM International Conference on Data Mining (SDM)*, 2010. Acceptance rate: 82/351 (23%)
7. **Pavel P. Kuksa**, Yanjun Qi, Bing Bai, Ronan Collobert, Jason Weston, Vladimir Pavlovic, and Xia Ning. Semi-supervised abstraction-augmented string kernel for multi-level bio-relation extraction. In *European Conference on Machine Learning (ECML)*, 2010. Acceptance rate: 106/658 (16%).
8. **Pavel P. Kuksa** and Vladimir Pavlovic. Spatial representation for efficient sequence classification. In *International Conference on Pattern Recognition (ICPR)*, 2010. Acceptance rate: 385/2140 oral (18%).

9. **Pavel Kuksa**, Pai-Hsi Huang, and Vladimir Pavlovic. Scalable algorithms for string kernels with inexact matching. In *Neural Information Processing Systems (NIPS)*, 2008. Spotlight Presentation. Acceptance rate: 123/1022 (12%)
10. **Pavel P. Kuksa** and Vladimir Pavlovic. Efficient motif finding algorithms for large-alphabet inputs. In *International Conference on Research in Computational Molecular Biology (RECOMB)*, 2010. Poster presentation.
11. **Pavel P. Kuksa** and Vladimir Pavlovic. Efficient motif finding algorithms for large-alphabet inputs. In *BIOKDD*, 2010. Acceptance rate: 7/29 regular (24%).
12. Yanjun Qi, Ronan Collobert, **Pavel Kuksa**, Koray Kavukcuoglu, and Jason Weston. Combining labeled and unlabeled data with word-class distribution learning. In *Proceeding of the 18th ACM Conference on Information and Knowledge Management (CIKM)*, pages 1737–1740, 2009. Acceptance rate: (123+171)/847 (20% short paper).
13. Yanjun Qi, **Pavel P. Kuksa**, Ronan Collobert, Kunihiko Sadamasa, Koray Kavukcuoglu, and Jason Weston. Semi-supervised sequence labeling with self-learned features. In *Proc. International Conference on Data Mining (ICDM)*. IEEE, 2009. Acceptance rate: 8.9% regular (70/786).
14. **Pavel Kuksa** and Vladimir Pavlovic. Fast motif selection for biological sequences. In *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2009. Acceptance rate: (44+37)/233 (35%).
15. **Pavel Kuksa** and Vladimir Pavlovic. Efficient alignment-free barcode analytics. In *Third International Barcode of Life Conference*, 2009.
16. **Pavel Kuksa**, Pai-Hsi Huang, and Vladimir Pavlovic. Fast and accurate multi-class protein fold recognition with spatial sample kernels. In *Computational Systems Bioinformatics (CSB)*, pages 133–143, 2008. Acceptance rate: 30/135 (22%).
17. **Pavel Kuksa**, Pai-Hsi Huang, and Vladimir Pavlovic. Fast protein homology and fold detection with sparse spatial sample kernels. In *19th International Conference on Pattern Recognition (ICPR)*, 2008. Acceptance rate: 18% (oral). *Best paper nominee*.
18. **Pavel Kuksa**, Pai-Hsi Huang, and Vladimir Pavlovic. A fast, semi-supervised learning method for protein sequence classification. In *8th International Workshop on Data Mining in Bioinformatics (BIOKDD)*, pages 29–37, 2008. Acceptance rate: 8/25 (32%).
19. **Pavel Kuksa**, Pai-Hsi Huang, and Vladimir Pavlovic. On the role of local matching for efficient semi-supervised protein sequence classification. In *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2008. Acceptance rate: 38/156 (24%).
20. **Pavel Kuksa** and Vladimir Pavlovic. Fast barcode-based species identification using string kernels. In *Second International Barcode of Life Conference*, 2007. Acceptance rate: 30% (oral).
21. **Pavel Kuksa** and Vladimir Pavlovic. Fast kernel methods for SVM sequence classifiers. In *Workshop on Algorithms in Bioinformatics (WABI)*, pages 228–239, 2007. Acceptance rate: 37/131 (28%).

- **Peer-reviewed Workshop Papers**

1. **Pavel P. Kuksa**. Efficient time series classification with multivariate similarity kernels. In *NYAS Machine Learning Symposium*, 2012. *Oral presentation*. Acceptance rate: 15/62.
2. **Pavel P. Kuksa**. 2d similarity kernels and representations for sequence data. Snowbird Learning Workshop, Snowbird, Utah, April 2012, 2012.
3. **Pavel P. Kuksa**. Using string kernels to predict gene expression. Snowbird Learning Workshop, Snowbird, Utah, April 2012, 2012.
4. **Pavel Kuksa** and Vladimir Pavlovic. Efficient evaluation of large sequence kernels. In *NYAS Machine Learning Symposium*, 2011.
5. **Pavel Kuksa** and Vladimir Pavlovic. Efficient sequence classification with spatial representations. In *Snowbird Learning Workshop*, April 2010. Oral presentation [28/69].
6. Vladimir Pavlovic and **Pavel Kuksa**. Large scale sequence analytics. In *Center for Dynamic Data Analytics (CDDA) Workshop (January 25-26, 2010)*, 2010.
7. **Pavel Kuksa** and Vladimir Pavlovic. Efficient discovery of common patterns in sequences. In *Snowbird Learning Workshop*, Clearwater, Florida, April 13-16 2009, 2009.
8. Jason Weston, Ronan Collobert, Frederic Ratle, Hossein Mobahi, **Pavel Kuksa**, and Koray Kavukcuoglu. Deep learning via semi-supervised embedding. In *ICML 2009 Workshop on Learning Feature Hierarchies*, 2009.
9. **Pavel Kuksa** and Vladimir Pavlovic. String kernel-based species identification using DNA barcodes. In *Joint Molecular Biosciences Symposium*, Feb 29, 2008, 2008.
10. **Pavel Kuksa**, Pai-Hsi Huang, and Vladimir Pavlovic. High performance sequence classification with novel spatial sample embedding. In *3rd Annual Machine Learning Symposium*, NY, Oct 10, 2008, 2008.
11. **Pavel Kuksa**, Pai-Hsi Huang, and Vladimir Pavlovic. Spatially-constrained sample kernel for sequence classification. In *Snowbird Learning Workshop*, Utah, April 1-4, 2008, 2008.
12. **Pavel Kuksa** and Vladimir Pavlovic. Kernel methods for DNA barcoding. In *Snowbird Learning Workshop*, San Juan, Puerto Rico, March 2007, 2007.

- **Peer-reviewed Conference Abstracts**

1. Matei Ionita*, **Pavel P. Kuksa***, Yuk Yee Leung, Kaylyn Clark, Otto Valladares, Li-San Wang. Bayesian Tissue Significance (BTS) scores for functional evaluation of GWAS results. In *American Society of Human Genetics Annual Meeting (ASHG)*, 2022.
2. J. Cifello*, **Pavel P. Kuksa***, L-S. Wang, Y.Y. Leung. Seamless integrative pipeline for QTL datasets enhance the discovery of putative causal variants for Alzheimer's Disease. In *American Society of Human Genetics Annual Meeting (ASHG)*, 2022.

3. YY Leung, B. Kunkle, B. Vardarajan, E. Marcora, E. Blue, **P. Kuksa**, R. M. Cores, A. DeStefano, R. Mayeux, L. Farrer, A. Goate, L.-S. Wang, G. Schellenberg, Alzheimer's Disease Sequencing Project. Integrative analyses to identify plausible target genes on Alzheimer's Disease genes with genetic evidence. In *American Society of Human Genetics Annual Meeting (ASHG)*, 2022. (**top 10% Reviewer's Choice**)
4. Jeffrey Cifello*, **Pavel P. Kuksa***, Li-San Wang, Yuk Yee Leung. Seamless integrative pipeline for QTL datasets enhance the discovery of putative causal variants for Alzheimer's Disease. In *MidAtlantic Bioinformatics Conference (MABC)*, 2022.
5. Matei Ionita, **Pavel P. Kuksa**, Chixiang Chen, Yuk Yee Leung, and Li-San Wang. Dissecting multiple-signal GWAS loci and their regulatory roles in Alzheimer's Disease. In *American Society of Human Genetics Annual Meeting (ASHG)*, 2021.
6. Emily Greenfest-Allen, Conor Klamann, Elizabeth E. Mlynarski, **Pavel P. Kuksa**, Prabhakaran Gangadharan, Amanda B Kuzma, Otto Valladares, Yuk Yee Leung, Christian J Stoeckert Jr., and Li-San Wang. Introducing the NIAGADS Alzheimer's GenomicsDB API: a toolkit for remote exploration of Alzheimer's disease genetics. In *Alzheimer's Association International Conference (AAIC)*, 2021.
7. Yuk Yee Leung, **Pavel P. Kuksa**, Yi-Fan Chou, Chia-Lun Liu, Wei Fu, Liming Qu, Yi Zhao, Zivadin Katanic, Amanda B Kuzma, Pei chuan Ho, Kai-Teh Tzeng, Otto Valladares, Shin-Yi Chou, Adam C Naj, and Gerald D. Schellenberg. Characterization of regulatory roles of genetics signals curated from >200 GWA studies in Alzheimer's Disease Variant Portal (ADVP). In *Alzheimer's Association International Conference (AAIC)*, 2021.
8. **Pavel. P. Kuksa**, P. Gangadharan, L. Kleidermacher, E. Greenfest-Allen, C.-Y. Lee, H.-J. Lin, Y.-F. Chou, Z. Katanic, O. Valladares, Y.Y. Leung, and L.-S. Wang. FILER: Harmonized, scalable functional genomics repository and api. In *ENCODE 2020 Research Applications & Users meeting*, 2020. **Oral presentation**. [<https://www.youtube.com/watch?v=dG0-iziv7o4>].
9. **Kuksa Pavel P.**, Lee C.-Y., Gangadharan P., Valladares O., Leung YY, and Wang L.-S. Spark-inferno: next-generation high-throughput pipeline for inferring molecular mechanisms of non-coding genetic associations. In *Program in Quantitative Genomics (PQG)*, 2020.
10. **Pavel. P. Kuksa**, P. Gangadharan, L. Kleidermacher, E. Greenfest-Allen, C.-Y. Lee, H.-J. Lin, Y.-F. Chou, Z. Katanic, O. Valladares, Y.Y. Leung, and L.-S. Wang. FILER: Integrated, large-scale functional genomics repository. In *American Society of Human Genetics Annual Meeting (ASHG)*, 2020.
11. Yuk Yee Leung, **Pavel P. Kuksa**, Chia-Lun Liu, Wei Fu, Liming Qu, Yi Zhao, Zivadin Katanic, Amanda B Kuzma, Pei chuan Ho, Kai-Teh Tzeng, Otto Valladares, Shin-Yi Chou, Adam C Naj, Gerard D Schellenberg, and Li-San Wang. Alzheimer's disease variant portal (ADVP): harmonized genetics data and evidence collection for Alzheimer's disease. In *Alzheimer's Association International Conference (AAIC)*, 2020.
12. Lauren Kleidermacher, **Pavel P. Kuksa**, Chien-Yueh Lee, Alexandre Amlie-Wolf, Gerard Schellenberg, Li-San Wang, and Yuk Yee Leung. Genome-wide profiling of the noncoding regulatory mechanisms in Alzheimer's Disease. In *Alzheimer's Association International Conference (AAIC)*, 2020.

13. **Pavel P. Kuksa**, Prabhakaran Gangadharan, Chien-Yueh Lee, Yi-Fan Chou, Emily Greenfest-Allen, Han-Jen Lin, Ž. Katanić, Otto Valladares, Yuk Yee Leung, Li-San Wang GADB: Large-scale, curated Functional Genomics Annotation Database In *American Society of Human Genetics Annual Meeting (ASHG)*, 2019.
14. C.-Y. Lee, **Pavel P. Kuksa**, A. Amlie-Wolf, E.E. Mlynarski, Y.-F. Chou, H.-J. Lin, E. Greenfest-Allen, Z. Katanic, O. Valladares, A. Kuzma, A. Naj, G.D. Schellenberg, Y.Y. Leung, L.-S. Wang, Alzheimer's Disease Sequencing Project INFERNO2: Scalable Spark-based framework for inferring dysregulated enhancer and noncoding RNAs for WGS and GWAS data In *American Society of Human Genetics Annual Meeting (ASHG)*, 2019.
15. Y.Y. Leung, **Pavel P. Kuksa**, C.-Y. Lee, Y.-F. Chou, A. Amlie-Wolf, G.D. Schellenberg, L.-S. Wang Non-coding regulatory landscape of Alzheimer's disease variants using GWAS of 63,926 individuals In *American Society of Human Genetics Annual Meeting (ASHG)*, 2019.
16. **Pavel P. Kuksa**, A. Amlie-Wolf, Y.-C. Hwang, B. D. Gregory, L.-S. Wang. Hi-C-based characterization of the landscape of physically interacting regions and interaction mechanisms across six human cell lines using HiPPIE2. In *American Society of Human Genetics Annual Meeting (ASHG)*, 2018.
17. E.E. Mlynarski, A. Amlie-Wolf, **Pavel P. Kuksa**, O. Valladares, G.D. Schellenberg, L.-S. Wang. SV-INFERNO: A Spark based pipeline for INFERring the molecular mechanisms of NONcoding structural variants. In *American Society of Human Genetics Annual Meeting (ASHG)*, 2018.
18. A. Amlie-Wolf, L. Qu, E.E. Mlynarski, **Pavel P. Kuksa**, Y.Y. Leung, C.D. Brown, G.D. Schellenberg, L.S. Wang. Inferring enhancer and noncoding RNA dysregulation underlying 2,419 UK Biobank phenotypes. In *American Society of Human Genetics Annual Meeting (ASHG)*, 2018. (**Platform talk**)
19. Y.Y. Leung*, **Pavel P. Kuksa***, A. Amlie-Wolf, L.-S. Wang. The landscape of short RNAs in human cell types and tissues. In *American Society of Human Genetics Annual Meeting (ASHG)*, 2017. (**top 10% Reviewer's Choice**)
20. **Pavel P. Kuksa**, Y.Y. Leung, A. Amlie-Wolf, O. Valladares, L.-S. Wang. DASHR 2.0: Database of small non-coding RNAs in normal human tissues and cell types. In *American Society of Human Genetics Annual Meeting (ASHG)*, 2017.
21. A. Amlie-Wolf, M. Tang, **Pavel P. Kuksa**, Y.Y. Leung, B. Slaff, J. King, B. Dombroski, G.D. Schellenberg, L.-S. Wang. INFERNO – INFERring the molecular mechanisms of NONcoding genetic variants. In *American Society of Human Genetics Annual Meeting (ASHG)*, 2016.
22. Y.Y. Leung, **Pavel P. Kuksa**, A. Amlie-Wolf, L.-S. Wang. The landscape of regulatory post-transcriptionally derived small non-coding RNAs in the human transcriptome. In *American Society of Human Genetics Annual Meeting (ASHG)*, 2016.
23. Y.-C. Hwang*, **P. P. Kuksa***, B. D. Gregory, and L.-S. Wang. Identifying the transcription factors mediating enhancer–target gene regulation in the human genome. In *American Society of Human Genetics Annual Meeting (ASHG)*, 2015. (**Platform talk**)

24. **P. P. Kuksa**, Y. Y. Leung, A. Amlie-Wolf, B. D. Gregory, and L.-S. Wang. SPAR: Sequencing-based pipeline for annotating novel small non-coding RNAs. In *American Society of Human Genetics Annual Meeting (ASHG)*, 2015.
25. Y. Y. Leung*, **P. P. Kuksa***, A. Amlie-Wolf, O. Valladares, B. D. Gregory, and L.-S. Wang. DASHR - database of small human non-coding RNAs. In *American Society of Human Genetics Annual Meeting (ASHG)*, 2015.
26. Mitchell Tang, Adam Naj, Li-San Wang, Gerald Schellenberg, Christian Kramer, George Xu, Michele Hawk, Yih-Chii Hwang, Chiao-Feng Lin, **Pavel P. Kuksa**, Weixin Wang, Beth Dombroski. Prediction of Late-Onset Alzheimer's Disease Associated Enhancer Elements In *Alzheimer's Association International Conference (AAIC)*, 2015.

- **Technical reports**

1. **Pavel P. Kuksa**. Efficient multivariate sequence classification. *CoRR abs/1409.8211*, 2013.
2. **Pavel P. Kuksa**, Imdadullah Khan, and Vladimir Pavlovic. Generalized similarity kernels for efficient sequence classification. Technical Report RU-DCS-TR684, Rutgers University, 2011.
3. Ronan Collobert, Jason Weston, Leon Bottou, Michael Karlen, Koray Kavukcuoglu, and **Pavel Kuksa**. Natural language processing (almost) from scratch. *arXiv:1103.0398v1*, 2011.
4. Robert S. Moore, Richard Howard, **Pavel Kuksa**, and Richard P. Martin. A geometric approach to device-free motion localization using signal strength. Technical Report DCS-TR-674, Rutgers University, September 2010.
5. **Pavel Kuksa** and Vladimir Pavlovic. Efficient discovery of common patterns in sequences over large alphabets. Technical Report 2009-15, DIMACS, June 2009.
6. **Pavel Kuksa**, Pai-Hsi Huang, and Vladimir Pavlovic. Kernel Methods and Algorithms for General Sequence Analysis. Technical Report DCS-TR-630, Rutgers University, 2008.
7. Pai-Hsi Huang, **Pavel Kuksa**, and Vladimir Pavlovic. Fast and accurate semi-supervised protein homology detection with large uncurated sequence databases. Technical Report RU-DCS-TR634, Rutgers University, 2008.

Patents

1. Yanjun Qi, Xia Ning, **Pavel P. Kuksa**, Bing Bai. Systems and Methods for Semi-Supervised Relationship Extraction. US 8874432 B2. Issued Oct 28, 2014.
2. Martin Renqiang Min, **Pavel P. Kuksa**, and Xia Ning. High-Order Semi-Restricted Boltzmann Machines and Deep Models for Accurate Peptide-MHC Binding Prediction and Collaborative Filtering. US 20150278441 A1. Filed 10/10/2014.
3. **Pavel P. Kuksa** and Martin Renqiang Min. High-Order Sequence Kernel Methods for Peptide Analysis. US 20160232281 A1. Filed 10/10/2014.

Invited Lectures and Oral Presentations

- FILER: Harmonized, scalable Functional Genomics Repository and API. ENCODE 2020: Research Applications and Users Meeting. October 1, 2020. (talk)
- SparkINFERNO: Scalable, harmonized big genomic and genetic data analysis. ADGC Annual Meeting. Philadelphia, February 25, 2020. (talk)
- Identifying the transcription factors mediating enhancer-target gene regulation in the human genome. Spatial and Functional Genomics Symposium. University of Pennsylvania, January 08, 2018. (talk)
- HiPR: High-throughput sequencing-based Probabilistic RNA structure inference. RNA club. University of Pennsylvania, September 19, 2017. (talk)
- Multivariate and generalized similarity kernels for sequence analysis. New Jersey Institute of Technology. March 7, 2013 (invited lecture)
- String kernel-based Species Identification Using DNA Barcodes. Rutgers-UMDNJ Joint Molecular Biosciences Symposium. Feb. 29, 2008 (invited lecture)
- Kernel Methods for DNA Barcoding. Rutgers Bioinformatics meeting, Nov. 10, 2006 (invited talk)
- Efficient time series classification with multivariate similarity kernels. NYAS Machine Learning Symposium. New York, NY, Oct. 19, 2012 (oral presentation)
- Efficient evaluation of large sequence kernels. KDD. Aug. 14, 2012 (oral presentation)
- Fast and accurate multi-class protein fold recognition with spatial sample kernels. Computational Systems Bioinformatics. Stanford University, Aug. 27, 2008 (oral presentation)
- Fast protein homology and fold detection with sparse spatial sample kernels. International Conference on Pattern Recognition. Tampa, FL, Dec. 9, 2008 (oral presentation)
- A fast, semi-supervised learning method for protein sequence classification. International Workshop on Data Mining in Bioinformatics. Las Vegas, Aug. 24, 2008 (oral presentation)
- On the role of local matching for efficient semi-supervised protein sequence classification. BIBM. Philadelphia, Nov. 4, 2008 (oral presentation)
- Fast kernel methods for SVM sequence classifiers. WABI. University of Pennsylvania, Sept. 8, 2007 (oral presentation)
- Efficient motif finding algorithms for large-alphabet inputs. BIODKDD. Washington, DC, July 25, 2010 (oral presentation)
- Efficient sequence classification with spatial representations. Snowbird Learning Workshop, Utah, Apr. 8, 2010 (oral presentation)

Awards and Honors

- American Society of Human Genetics (ASHG) top 10% Reviewer's Choice poster abstract, 2017
- DIMACS Graduate Student Award, Center for Discrete Mathematics and Theoretical Computer Science, Jan 2010
- DIMACS Graduate Student Award, Center for Discrete Mathematics and Theoretical Computer Science, Jan 2009

- Best Paper award nomination at the International Conference on Pattern Recognition in Tampa, FL, 2008
- Graduate Fellowship Award, Rutgers University, 2005-2007
- President of Russia's Award for Outstanding Young Scientists, Ministry of Education and Science, Russia, 2004
- Scientific Council Scholarship Award, Bauman Moscow State Technical University (BMSTU), 2002-2003
- Award for Academic Excellence, Bauman Moscow State Technical University (BMSTU), 2001, 2002
- M.Sc. with Honors, 2004
- B.Sc. with Honors, 2002

Professional Services & Activities

- **Journal article refereeing**
 1. Nucleic Acids Research (NAR)
 2. Bioinformatics
 3. Nature Scientific Reports
 4. BMC Bioinformatics
 5. PLOS Computational Biology
 6. Briefings in Bioinformatics
 7. Frontiers in Genetics
 8. IEEE Transactions on Computational Biology and Bioinformatics (TCBB)
 9. IEEE Transactions on Language and Speech Processing (TLSP)
 10. IEEE Access
 11. Proceedings of the Royal Society
 12. Algorithmica
 13. Algorithms
 14. Neural Networks
 15. Neural Computation
 16. Neural Processing Letters
 17. Machine Learning
 18. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)
 19. IEEE Computational Intelligence Magazine (CIM)
 20. Journal of Machine Learning Research (JMLR)
 21. Evolutionary Bioinformatics
 22. Microorganisms
 23. Biomolecules
 24. Aquatic Biosystems
 25. Molecular Informatics
 26. Molecular Ecology Resources
 27. Molecular Sciences

28. Information Sciences

29. PLOS One

- **Conference Referee**

1. International Conference on Machine Learning (ICML)
2. IEEE International conference on Pattern Recognition (ICPR)
3. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)
4. International Conference on Computer Vision (ICCV)
5. NIPS Machine Learning in Computational Biology (MLCB)
6. IEEE International Conference on Bioinformatics and Biomedicine (BIBM)
7. International Symposium on Bioinformatics Research and Applications (ISBRA)
8. International Conference on Computational Advances in Bio and medical Sciences (ICCABS)

- **Program committee and organizer**

1. Funding proposal refereeing. Marsden Fund, Royal Society of New Zealand, 2021
2. Grant proposal review. NWO Dutch Research Council, 2015
3. Grant proposal review. European Coordinated Research on Long-term Challenges in Information and Communication Sciences & Technologies ERA-Net (CHIST-ERA), 2015
4. Arts and Sciences Council, The Perelman School of Medicine Art in Science Competition, University of Pennsylvania, 2014
5. Program committee, Biomedical Postdoctoral Research Symposium, University of Pennsylvania, 2014
6. Data Analysis Working Group (DAWG), International Consortium for Barcode of Life (CBOL), 2008-2011
7. Co-organizer, Yahoo Machine Learning seminar series, Rutgers University, 2010

- **Teaching**

1. Introduction to Bioinformatics (CS674), PhD level, Rutgers University
2. Discrete Structures II (CS206), Undergraduate, Rutgers University
3. Principles of Artificial Intelligence (CS530), PhD level, Rutgers University
4. Pattern Recognition (CS535), PhD level, Rutgers University
5. Machine Learning (CS536), PhD level, Rutgers University
6. Introduction to Artificial Intelligence (CS440), Undergraduate, Rutgers University
7. Computer Applications for Business (CS170), Undergraduate, Rutgers University
8. Systems modeling, Post-graduate level, Informatics and Control Systems Department, Bauman Moscow State Technical University (BMSTU)
9. Computer Systems Design, Graduate level, Informatics and Control Systems Department, Bauman Moscow State Technical University (BMSTU)

- **Service**

- Senior Fellow, Institute for Biomedical Informatics (IBI), University of Pennsylvania, Mar 2023–present
- Member, The American Society of Human Genetics, 2015–present

- Associate Fellow, Institute for Biomedical Informatics (IBI), University of Pennsylvania, Oct 2014–Mar 2023
- Treasurer/Secretary, Computer Science Graduate Student Society, Rutgers University, Spring 2006–Fall 2008
- Secretary, Rutgers Bioinformatics Club, 2006–2007
- Member of IEEE, 2005–present

Teaching Interests

- Bioinformatics, Data Mining, Biomedical Informatics, Machine Learning, Design and Analysis of Algorithms, Natural Language Processing, Pattern Recognition

Appendix A. Publications in Russian

1. Kouxa, Pavel. "Applications of Associative Memory and Associative Processors in Network Devices: A Survey." (Primenenie asociativnih ZU i processorov v setevih ustroystvah) Modern Information Technologies. Moscow: BMSTU Press, 2001. 20-24. ISBN 5-7038-2083-9
2. Kouxa, Pavel. "Data Clustering with Neural Networks." (Primenenie neironnih setei dlya klasterizatsii dannih) Modern Information Technologies. Moscow: BMSTU Press, 2001. 7-13. ISBN 5-7038-2083-9
3. Kouxa, Pavel. "Finding Logically Independent Operators Matrix." (Algoritm nahojdeniya matricy logicheskoi nesovmestivosti) Informatics and Control Systems in 21st century. Moscow: Elik+ Press, 2002. 195-97. ISBN 5-93991-010-6
4. Kouxa, Pavel. "Fuzzy Systems Simulation Software at Algorithmic Level." (Sistema modelirovaniya nechetkih system na algoritmicheskom urovne) Informatics and Control Systems in 21st century. Moscow, 2002. 201-04. ISBN 5-93991-010-6
5. Kouxa, Pavel. "Finding Logically Independent Operators for Parallel Computations Scheduling." (Vychislenie matritsy logicheskoi nesovmestivosti pri organizatsii parallel'nyh vychislenii) Young Scientists, Graduate and Undergraduate Students Conference. Moscow: BMSTU Press, 2003. 293-97. ISBN 5-7038-2347-1
6. Kouxa, Pavel. "Fuzzy Clustering Algorithm: Reducing Complexity." (Analiz algoritma nechetkoi klasterizatsii) Young Scientists, Graduate and Undergraduate Students Conference. Moscow: BMSTU Press, 2003. 249-53. ISBN 5-7038-2347-1
7. Kouxa, Pavel. "Fuzzy Systems Definition Language." (Yazyk opisaniya nechetkih sistem) Journal of Volzskiy Tatishev University, Information Science Series. 5 (2004): 92-97.
8. Kouxa, Pavel. "Linguistic Fuzzy Models Structure Identification and Optimization." (Struktural'naya identifikatsiya i optimizatsiya lingvisticheskikh nechetkih modelei) Modern Information Technologies. Moscow: Elik+, 2003. 7-16. ISBN 5-93991-014-9
9. Kouxa, Pavel. "Neuro-Fuzzy Models: Analysis and Applications." (Analiz neuro-nechetkih modelei) Journal of Volzskiy Tatishev University, Information Science Series.5 (2004): 69-76.
10. Kouxa, Pavel. "Software Implementation of Fuzzy Logic Controllers." (Programmnaia realizatsiya nechetkih kontrollerov) Modern Information Technologies. Moscow: Elik+, 2003. ISBN 5-93991-014-9
11. Kouxa, Pavel. "Type-2 Fuzzy Models." (Nechetkie lingvisticheskie modeli vtorogo roda) Journal of Volzskiy Tatishev University, Information Science Series.5 (2004): 114-20.
12. Kouxa, Pavel. "Type-2 Fuzzy Sets Algebra." (Issledovanie operatsii algebry nechetkih mnojestv vtorogo poryadka) Young Scientists, Graduate and Undergraduate Students Conference, Informatics and Control Systems. Moscow: BMSTU Press, 2003. 243-48. ISBN 5-7038-2347-1
13. Kouxa, Pavel. "Approximation Accuracy of Fuzzy Systems." (Obespechenie tochnosti v nechetkih sistemah) Young Scientists, Graduate and Undergraduate Students Conference. Informatics and Control Systems in 21st century. Moscow: BMSTU Press, 2004. 152-156. ISBN 5-7038-2646-4
14. Kouxa, Pavel. "Design Optimization of Non-Local Neuro-Fuzzy Models." (Optimizatsiya i sintez nelokal'nykh lingvisticheskikh neuro-nechetkih modelei) Inter-universities Scientific and Technical Conference on Modern Information Technologies. Moscow, 2004.

15. Kouxa, P. "Fuzzy Systems: Learning and Self-Organization Methods." Young Scientists, Graduate and Undergraduate Students Conference. Informatics and Control Systems in 21st century. Moscow: BMSTU Press, 2004. 148-52. ISBN 5-7038-2646-4
16. Kouxa Pavel, Schmakov E., Yahina I., Panushkin M. "A Survey on Supercomputing: Modern State and Perspectives." (Superkomputery: sostoyanie i perspektivy) Anniversary Inter-universities Scientific and Technical Conference on "Modern Information Technologies". Moscow: BMSTU Press, 2000. 61-68. ISBN 5-7038-1752-8
17. Timofeev V.V., Kouxa P. "Design and Identification of Fuzzy Models in Control and Systems Modeling." (Postroenie i identifikatsiya nechetkih modelei v prilozeniyah upravleniya i sistemnogo modelirovaniya) First International Conference "Aerospace Technologies." Moscow: BMSTU Press, 2004. ISBN 5-7038-2517-2