Sangseon Lee (이상선)

SNU Bioinformatics Institute
Seoul National University
Seoul, Korea
Phone. +82-10-2928-3740
E-mail. sangseon486@snu.ac.kr

EDUCATION

Seoul National University, Korea

- Ph.D., Computer Science and Engineering   Mar. 2014 – Feb. 2020
  (Advisor: Professor Sun Kim)

- B. S., Computer Science and Engineering (*Summa Cum Laude*)   Mar. 2010 – Aug. 2013

EXPERIENCE

Seoul National University, Korea

- Postdoctoral research fellow, SNU Bioinformatics Institute   Mar. 2020 – present
  (Advisor: Professor Sun Kim)

RESEARCH INTERESTS

My general research area is applied machine learning and information theory for interpretation of biological systems. Focuses on modeling interactions between biological elements of living organisms from DNA sequences to gene expression or epigenomes. Applications include subpath mining, sequence similarity measurement, interpretable deep learning model, and computational biology with multi-omics analysis.

RESEARCH EXPERIENCE

Graduate research in Bio & Health Informatics Lab, Seoul National University, Korea

< Pathway analysis for high-dimension & small sample-size data >
- Cancer subtype classification and modeling by pathway attention and propagation
  - Designing a pathway-based explainable deep learning model by graph convolutional network and attention mechanism
  - Predicting cancer subtypes using gene expression data and pathway information
  - Identifying subtype related transcription factors using network propagation algorithm
  - Analyzed RNA-seq dataset of five cancers with multiple subtypes
  - Paper accepted by *Bioinformatics* journal (SCI, IF=4.531)

- Comprehensive and critical evaluation of individualized pathway activity measurement tools on pan-cancer data
  - Suggesting five criteria for measuring usefulness of pathway activity inference tools
  - Analyzed 13 pathway activity inference tools on PAN-cancer data
  - Paper accepted by *Briefings in Bioinformatics* journal (SCI, IF=9.101)

- MIDAS: Mining differentially activated subpaths of KEGG pathways from multi-class RNA-seq data
  - Kinds of subgraph mining for identifying differentially activated subpaths among multi-class data
  - Handling multi-class data by statistical test & greedy-based expansion technique
  - Paper accepted by *Methods* journal (SCI, IF=3.998)

< Deep Learning Models using Prior Knowledge >

- Learning Cell-Type-Specific Gene Regulation Mechanisms by Multi-Attention Based Deep Learning with Regulatory Latent Space
    - Designing multi-modal deep learning model for learning gene regulation mechanism
    - Exploring properties of each omics by different neural network structures (ex. CNN, RNN)
    - Elucidating epigenetic gene regulation mechanism by various attention layers
    - Paper submitted in *Frontiers in Genetics* journal

- Simultaneous Modelling of Hierarchy in GPCR Proteins with Deep Learning in a Single Metric Space
    - Learning hierarchy in GPCR proteins using a multi-task convolutional neural network
    - Utilizing two types of loss: softmax-loss (for inter-class) & center-loss (for intra-class)
    - Encapsulating biological properties (ex. Phylogeny, Motif) in embedded hidden features
    - Paper submitted in *European Conference on Computational Biology (ECCB) 2020* conference

< Information theoretic approach in sequence-type data >

- Ranked k-spectrum kernel for comparative and evolutionary comparison of exons, introns, and CpG islands
    - Suggesting a sequence similarity kernel based on rank and evolutionary information
    - Reconstructing phylogenetic trees of 10 mammalian species using exon, intron, and CpG islands sequences
    - Finding relationship across three regions in terms of evolutionary information contents: exon > CpG island > intron
    - Paper accepted by *IEEE/ACM Transactions on Computational Biology and Bioinformatics* journal (SCIE, IF=2.896)

- SpliceHetero: An information theoretic approach for measuring spliceomic intratumor heterogeneity from bulk tumor RNA-seq
    - Measuring intratumor heterogeneity in transcriptome level by handling spliceome
    - Utilizing Shannon entropy and Jensen-Shannon Divergence to capture heterogeneity
    - Paper accepted by *PloS ONE* journal (SCI, IF=2.776)

< Application of computational algorithms on epigenomic data >

- BioVLAB-mCpG-SNP-EXPRESS: A system for multi-level and multi-perspective analysis and exploration of DNA methylation, sequence variation (SNPs), and gene expression from multi-omics data
    - Platform for analyzing complex relationships between omics data
    - Guiding interpretation of results by multi-level and multi-perspective panels
    - Paper accepted by *Methods* journal (SCIE, IF=3.802)

- PRISM: Methylation Pattern-based, Reference-free Inference of Subclonal Makeup
    - Inferring the composition of epigenetically distinct subclones of a tumor from methylation patterns
    - Proofreading of methylation pattern based on DNMT1-like HMM
    - Inferencing epigenetic subclonal information by k-mixture decomposition with EM algorithm
    - Paper accepted by *ISMB 2019* conference, *Bioinformatics* journal (SCI, IF=4.531)

TEACHING EXPERIENCE

Guest lecturer: 생물정보학을 위한 IT기초, 2020 Spring, Seoul National University, Korea

Guest lecturer: 기계학습의 이해, 2019 Oct, Big Data Academy, Seoul National University, Korea

Guest lecturer: 기계학습의 이해, 2018 Jul, Big Data Academy, Seoul National University, Korea

Guest lecturer: Deep Learning for Bioinformatics, 2018 Feb, KSBi-BIML 2018 (생물정보학&머신러닝/인공지능 워크샵), Korea

PUBLICATIONS

Research Highlights

2 papers published/accepted in *Bioinformatics* (The flagship journal in the bioinformatics field)

1 paper presented at *Intelligent Systems for Molecular Biology (ISMB)* (한국정보과학회 선정 최우수 학술대회)

International Journal (*: equal contributors)

Lee S, Lim S, Lee T, Sung I, Kim S. Cancer subtype classification and modeling by pathway attention and propagation. ***Bioinformatics*** (SCI, IF=4.531), accepted.

Lim S, **Lee S**, Jung I, Rhee S, Kim S. Comprehensive and critical evaluation of individualized pathway activity measurement tools on pan-cancer data. ***Briefings in Bioinformatics*** (SCI, IF=9.101), 2020 Jan; 21(1):36-46.

Hur B, Kang D, **Lee S**, Moon JH, Lee G, Kim S. Venn-diaNet: Venn diagram based network propagation analysis framework for computing multiple biological experiments. ***BMC Bioinformatics*** (SCIE, IF=2.511), 2019 Dec 27; 20(23):1-12.

Kang D, Ahn H, **Lee S**, Lee CJ, Hur J, Jung W, Kim S. StressGenePred: a twin prediction model architecture for classifying the stress types of samples and discovering stress-related genes in Arabidopsis. ***BMC Genomics*** (SCIE, IF=3.501), 2019 Dec 20; 20(11):949.

Kim MS, **Lee S**, Lim S, Kim S. SpliceHetero: An information theoretic approach for measuring spliceomic intratumor heterogeneity from bulk tumor RNA-seq. ***PloS ONE*** (SCI, IF=2.776), 2019 Oct 23; 14(10).

Lee S, Lee T, Noh YK, Kim S. Ranked k-spectrum kernel for comparative and evolutionary comparison of exons, introns, and CpG islands. ***IEEE/ACM Transactions on Computational Biology and Bioinformatics*** (SCIE, IF=2.896), 2019 Sep 3, in press.

Lee DH. **Lee S**, Kim S. PRISM: Methylation Pattern-based, Reference-free Inference of Subclonal Makeup. ***Bioinformatics*** (SCI, IF=4.531), 2019 Jul 5; 35(14):i520-i529.

Lee CJ, Kang D, **Lee S**, Lee S, Kang J, Kim S. In silico experiment system for testing hypothesis on gene functions using three condition specific biological networks. ***Methods*** (SCI, IF=2.048), 2018 May 25; 145:10-15.

Lee S, Park Y, Kim S. MIDAS: Mining differentially activated subpaths of KEGG pathways from multi-class RNA-seq data. ***Methods*** (SCI, IF=3.998), 2017 Jul 15; 124:13-24.

Moon JH, Lim S, Jo K, **Lee S**, Seo S, Kim S, PINTnet: construction of condition-specic pathway interaction network by computing shortest paths on weighted PPI. ***BMC Systems Biology*** (SCI, IF=2.05), 2017 Mar 14; 11(2):15.

Chae H, **Lee S**, Nephew KP, Kim S, Subtype-specific CpG island shore methylation and mutation patterns in 30 breast cancer cell lines. ***BMC Systems Biology*** (SCI, IF=2.303), 2016 Dec 23; 10(4):116.

Chae H, **Lee S**, Seo S, Jung D, Chang H, Nephew KP, Kim S. BioVLAB-mCpG-SNP-EXPRESS: A system for multi-level and multi-perspective analysis and exploration of DNA methylation, sequence variation (SNPs), and gene expression from multi-omics data. ***Methods*** (SCI, IF=3.802), 2016 Jul 28; 111:64-71.

Jeong HM, **Lee S**, Chae H, Kim R, Kwon MJ, Oh E, Choi YL, Kim S, Shin YK. Efficiency of methylated DNA immunoprecipitation bisulphite sequencing for whole-genome DNA methylation analysis. ***Epigenomics*** (SCIE, IF=4.541), 2016 Jun 8; 8(8):1061-1077.

International Conference

Lee T, **Lee S**, Kang M, Sun Kim. Simultaneous Modelling of Hierarchy in GPCR Proteins with Deep Learning in a Single Metric Space. In preparation for ***European Conference on Computational Biology (ECCB) 2020***.

Hur B, Kang D, **Lee S**, Moon JH, Lee G, Kim S. Venn-diaNet: Venn diagram based network propagation analysis framework for computing multiple biological experiments. ***30th International Conference on Genome Informatics (GIW)*** 2019, Dec 9-11, Sydney, Australia.

Lee DH. **Lee S**, Kim S. PRISM: Methylation Pattern-based, Reference-free Inference of Subclonal Makeup. ***27th conference on Intelligent Systems for Molecular Biology (ISMB)*** 2019, July 21-25, Basel, Switzerland.

Kang D, Ahn H, **Lee S**, Lee CJ, Hur J, Jung W, Kim S. Identifying stress-related genes and predicting stress types in Arabidopsis using logical correlation layer and CMCL loss through time-series data. ***IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*** 2018, Dec 3-6, Madrid, Spain.

Lee S, Moon JH, Park Y, Kim S, Flow maximization analysis of cell cycle pathway activation status in breast cancer subtypes, ***IEEE International Conference on Big Data and Smart Computing (BigComp)*** 2017, Feb 13-16, Jeju, Korea.

Moon JH, Lim S, Jo K, **Lee S**, Seo S, Kim S, PINTnet: construction of condition-specic pathway interaction network by computing shortest paths on weighted PPI. ***15th Asia Pacific Bioinformatics Conference (APBC)*** 2017, Jan 16-18, Shenzhen, China.

Chae H, **Lee S**, Nephew KP, Kim S, Subtype-specific CpG island shore methylation and mutation patterns in 30 breast cancer cell lines. ***27th International Conference on Genome Informatics (GIW)*** 2016, Oct 3-5, Shanghai, China.

In Progress

Kang M*, **Lee S***, Lee D, Kim S. Learning Cell-Type-Specific Gene Regulation Mechanisms by Multi-Attention Based Deep Learning with Regulatory Latent Space. In preparation for ***Frontiers in Genetics***.

Moon JH, **Lee S**, Hur B, Kim S. MLDEG: A network-based ensemble model to identify differentially expressed genes consistently. In preparation for ***IEEE/ACM Transactions on Computational Biology and Bioinformatics***.

Lee T, **Lee S**, Kang M, Sun Kim. Simultaneous Modelling of Hierarchy in GPCR Proteins with Deep Learning in a Single Metric Space. In preparation for ***European Conference on Computational Biology (ECCB) 2020***.

Kim MS, **Lee S**, Lim S, Lee DY, Kim S. AttentiveCancer: an attention framework using subnetwork level representation of transcriptome for predicting lymph node metastasis in early oral cancer. In preparation for ***European Conference on Computational Biology (ECCB) 2020***.

GRANTS / AWARDS

Best Paper Award (Bronze), 한국정보과학회 인공지능소사이어티 The AI KOREA 2019, Jul 26, 2019.

Domestic Scholarship, 관정 이종환 교육재단, 2017-2018.

Best Paper Award on Bioinformatics, IEEE International Conference on Big Data and Smart Computing (BigComp) 2017, Feb 16, 2017.