Mohammed El-Kebir
Assistant Professor

Address: Department of Computer Science
University of Illinois Urbana-Champaign
Urbana, IL 61801, USA
Member: AAAS, ACM, IEEE, ISCB, SBME

Phone: +1 217-300-7369
E-mail: melkebir@illinois.edu
WWW: https://www.el-kebir.net
GitHub: https://github.com/elkebir-group

Research interests

Combinatorial optimization and its applications to cancer and infection genomics, integer linear programming, exact algorithms, graph algorithms, cancer phylogenetics, intra-tumor heterogeneity, biological network analysis.

Education

2010–2015	**Centrum Wiskunde & Informatica/VU University Amsterdam, The Netherlands.** Ph.D. in Computer Science and Bioinformatics Thesis: Networks, modules and breeding schedules: Applications of Combinatorial Optimization to Computational Biology Advisors: Drs. G. W. Klau and J. Heringa
2006–2010	**VU University Amsterdam, The Netherlands.** M.Sc. in Bioinformatics, *cum laude*. Thesis: Modeling Tuberculosis in Lung and Central Nervous System Advisors: Drs. D. Kirschner, M. van der Kuip, A. M. van Furth
2007–2009	**Eindhoven University of Technology, The Netherlands.** M.Sc. in Computer Science and Engineering, *cum laude*. Thesis: Crossing Schedule Optimization Advisors: Drs. M. T. de Berg and J. B. Buntjer
2003–2006	**Eindhoven University of Technology, The Netherlands.** B.Sc. in Computer Science and Engineering.

Honors and Awards

2021	○ National Science Foundation's CAREER Award
	○ Listed in "List of Teachers Ranked as Excellent by Their Students" for CS 598MEB: Computational Cancer Genomics (Spring 2019, Spring 2020, Spring 2021)
2019	○ Collins Scholar, Grainger College of Engineering, Academy for Excellence in Engineering Education, UIUC
2015	○ BioSB Young Investigator Award for best Ph.D. thesis in computational biology in the Netherlands
	○ ISCB ISMB/ECCB Travel Fellowship

2014	○ Honorable Mention for Best Paper Award (BioVis 2014)
2012	○ ISCB-RECOMB Travel Fellowship
2009	○ VU FEW Free mover grant and KNCV Tuberculosis Foundation grant

Positions

2018–present	**Department of Computer Science, University of Illinois at Urbana-Champaign, Urbana, IL, USA.**
	Assistant Professor. Affiliate appointments in National Center of Supercomputing Applications, Department of Electrical and Computer Engineering, Institute of Genomic Biology (Infection Genomics of One Health [IGOH] and Center for Artificial Intelligence and Modeling [CAIM] themes).
2016–2017	**Department of Computer Science, Princeton University, Princeton, NJ, USA.**
	Postdoctoral research associate.
2014–2016	**Department of Computer Science and Center for Computational Molecular Biology, Brown University, Providence, RI, USA.**
	Postdoctoral research associate.

Research Support

2021	○ 'CAREER: Algorithms for Comprehensive and Cost-effective Cancer Phylogeny Inference from Multi-omics Single-cell Sequencing Data'. *National Science Foundation (NSF 2046488), CISE Directorate, Division of Computing and Communication Foundations*. PI: M. El-Kebir. Budget: $500,000. Award period: 04/01/2021–03/31/2026.
	○ 'Algorithms for Single-cell Cancer Phylogenetics'. *Greg Gulick Honorary Research Awards*. PI: M. El-Kebir. Budget: $25,000 (AWS credits). Award period: 09/01/2021–10/31/2023.
	○ 'Workflows and Tools for Visualizing Tumor Phylogenies in Metastatic Breast Cancer'. *Cancer Center at Illinois (CCIL)*. Seed funding program. PI: M. El-Kebir. co-PI: C. Blatti (Research Scientist, NCSA, UIUC), Z. Madak-Erdogan (Assoc. Professor, FSHN, UIUC). Budget: $198,905. Award period: 07/01/2021–06/30/2023.
	○ 'Mayo Clinic and Illinois Alliance Fellowship for Technology-Based Healthcare Research'. *Interdisciplinary Health Science Institute*, University of Illinois at Urbana-Champaign. PIs: N. Chia (Assoc. Professor, Mayo), M. El-Kebir. Budget: $40,000. Award period: 01/16/2021–01/15/2023.
2020	○ 'RAPID: Deciphering Within-host Diversity and Multi-strain Infection in COVID-19'. *National Science Foundation (NSF 2027669), CISE Directorate, Division of Computing and Communication Foundations*. PI: M. El-Kebir. co-PI: J. Peng (Assist. Professor, CS, UIUC). Budget: $100,000. Award period: 05/15/2020–05/14/2021.
2019	○ 'CRII: AF: Towards an Accurate and Complete Characterization of the Solution Space in Phylogeny Estimation from Mixed Samples'. *National Science Foundation (NSF 1850502), CISE Directorate, Division of Computing and Communication Foundations*. PI: M. El-Kebir. Budget: $174,999. Award period: 06/15/2019–05/30/2021.
	○ 'Mayo Clinic and Illinois Alliance Fellowship for Technology-Based Healthcare Research'. *Interdisciplinary Health Science Institute*, University of Illinois at Urbana-Champaign. PIs: N. Chia (Assoc. Professor, Mayo), S. Koyejo (Assist. Professor, CS, UIUC), M. El-Kebir. Budget: $40,000. Award period: 08/16/2019–08/15/2021.

○ 'Algorithms for Experimental Study Design in Cancer Genomics', *Center for Computational Biotechnology and Genomic Medicine (CCBGM)*. PI: <u>M. El-Kebir</u>. co-PIs: N. Chia (Assoc. Professor, Mayo), S. Koyejo (Assist. Professor, CS, UIUC). Budget: $120,000. Award period: 01/01/2019–12/31/2020.

Publications

Author order generally follows authorship convention in biology, where the first author carried out the majority of the research (joint first authorship is indicated by '*') and the corresponding author supervised and/or conceived the project (indicated by '†'). Advised students are indicated in boldface. Many of the papers listed below first appeared in a conference and later in a journal. Only the journal version (with an additional indication of the conference) is listed if the two versions are very similar in content.

Edited conference proceedings

2021   21st International Workshop on Algorithms in Bioinformatics, WABI 2021, August 2–4, 2021, Virtual Conference. Edited by Alessandra Carbone and <u>Mohammed El-Kebir</u>.

Peer-reviewed journal

2022   **Z. Lalani***, **G. Chu***, S. Hsu, **S. Kagawa**, **M. Xiang**, S. Zaccaria† and <u>M. El-Kebir</u>†. CNAViz: An interactive webtool for user-guided segmentation of tumor DNA sequencing data. *PLOS Computational Biology* 18, no. 10 (October 13, 2022): e1010614, 2022. Extended version of RECOMB Computational Cancer Biology (RECOMB-CCB) 2022 paper.

C. Zhang*, **P. Sashittal***, **M. Xiang**, **Y. Zhang**, **A. Kazi** and <u>M. El-Kebir</u>†. Accurate Identification of Transcription Regulatory Sequences and Genes in Coronaviruses. *Molecular Biology and Evolution* 39, no. 7, July 2022. Extended version of Research in Computational Molecular Biology (RECOMB) 2022 paper.

P. Sashittal, S. Zaccaria and <u>M. El-Kebir</u>†. Parsimonious Clone Tree Integration in Cancer. *Algorithms for Molecular Biology* 17, no. 1 (2022): 3. Extended version of Workshop on Algorithms in Bioinformatics (WABI) 2021 paper.

C. Oh*, **P. Sashittal***, A. Zhou, L. Wang, <u>M. El-Kebir</u>†, and Thanh H. Nguyen†. Design of SARS-CoV-2 Variant-Specific PCR Assays Considering Regional and Temporal Characteristics. *Applied and Environmental Microbiology*, March 14, 2022, e02289-21.

2021   **P. Sashittal**, **C. Zhang**, J. Peng and <u>M. El-Kebir</u>†. Jumper Enables Discontinuous Transcript Assembly in Coronaviruses. *Nature Communications* 12, no. 1 (December 2021): 6728.

G. Satas*, S. Zaccaria*, <u>M. El-Kebir</u>*,† and B.J. Raphael†. DeCiFering the elusive cancer cell fraction in tumor heterogeneity and evolution. *Cell Systems*, 12:1004–1018, 2021. Published online August 2021. Extended version of selected papers that appeared in Research in Computational Molecular Biology (RECOMB) 2021.

L.L. Weber*, **P. Sashittal*** and <u>M. El-Kebir</u>†. doubletD: detecting doublets in single-cell DNA sequencing data. *Bioinformatics*. 37(Supplement_1):i214-i221, 2021. Special issue for Intelligent Systems for Molecular Biology/European Conference on Computational Biology (ISMB/ECCB) 2021.

L.L. Weber and <u>M. El-Kebir</u>†. Distinguishing linear and branched evolution given single-cell DNA sequencing data of tumors. *Algorithms for Molecular Biology*. 16(1):14, 2021. Extended version of Workshop on Algorithms in Bioinformatics (WABI) 2020 paper.

C. Zhang, M. El-Kebir[†] and I. Ochoa[†]. Moss enables high sensitivity single-nucleotide variant calling from multiple bulk DNA tumor samples. *Nature Communications*, 12(1):2204, 2021.

S.S. Patel, **A. Sandur**, M. El-Kebir, R.C. Gaba, L.B. Schook and K.M. Schachtschneider Transcriptional Profiling of Porcine HCC Xenografts Provides Insights Into Tumor Cell Microenvironment Signaling. *Frontiers in Genetics*, 12, 2021.

2020   **S. Christensen**, **J. Kim**, N. Chia, O. Koyejo and M. El-Kebir[†]. Detecting evolutionary patterns of cancers using consensus trees. *Bioinformatics*, 36:i684–i691, 2020. Special issue for European Conference on Computational Biology (ECCB) 2020.

L.L. Weber[*], **N. Aguse**[*], N. Chia and M. El-Kebir[†]. PhyDOSE: Design of Follow-up Single-cell Sequencing Experiments of Tumors. *PLOS Computational Biology*, 16(10):e1008240, 2020. Invited journal version of RECOMB Computational Cancer Biology (RECOMB-CCB) 2020 conference paper.

R. C. Gaba, L. Elkhadragy, F.E. Boas, S. Chaki, H.H. Chen, M. El-Kebir, K.D. Garcia, E.F. Giurini, G. Guzman, F.V. LoBianco, M.F. Neto, J.L. Newson, A. Qazi, M. Regan, L.A. und, R.M. Schwind, M.C. Stewart, F.M. Thomas, H.E. Whiteley, **J. Wu**, L.B. Schook and K.M. Schachtschneider. Development and comprehensive characterization of porcine hepatocellular carcinoma for translational liver cancer investigation. *Oncotarget* 11:2686--2701, 2020.

J. Wu and M. El-Kebir[†]. ClonArch: Visualizing the Spatial Clonal Architecture of Tumors. *Bioinformatics*, 36:i161–i168, 2020. Special issue for Intelligent Systems for Molecular Biology (ISMB) 2020.

P. Sashittal and M. El-Kebir[†]. Sampling and Summarizing Transmission Trees with Multi-strain Infections. *Bioinformatics*, 36:i362–i370, 2020. Special issue for Intelligent Systems for Molecular Biology (ISMB) 2020.

2019   **Y. Qi**, **D. Pradhan** and M. El-Kebir[†]. Implications of Non-uniqueness in Phylogenetic Deconvolution of Bulk DNA Samples of Tumors. *Algorithms for Molecular Biology*, 14:19, 2019. Extended version of RECOMB Comparative Genomics (RECOMB-CG) 2018 paper.

N. Aguse[*], **Y. Qi**[*] and M. El-Kebir[†]. Summarizing the Solution Space in Tumor Phylogeny Inference by Multiple Consensus Trees. *Bioinformatics*, 35(14):i408-i146, 2019. Special issue for Intelligent Systems for Molecular Biology/European Conference on Computational Biology (ISMB/ECCB) 2019.

2018   M. El-Kebir[†]. SPhyR: Tumor Phylogeny Estimation from Single-Cell Sequencing Data under Loss and Error. *Bioinformatics*, 34(17):i671-679, 2018. Special issue for European Conference on Computational Biology (ECCB) 2018.

P. Spohr, K. Dinkla, G.W. Klau and M. El-Kebir[†]. eXamine: Visualizing annotated networks in Cytoscape. *F1000Research*, 7, 519, 2018.

S. Zaccaria[*], M. El-Kebir[*], G.W. Klau, B.J. Raphael. Phylogenetic Copy-Number Factorization of Multiple Tumor Samples. *Journal of Computational Biology*, 27(0), 2018. Extended version of Research in Computational Molecular Biology (RECOMB) 2017 paper.

M. El-Kebir, G. Satas and B.J. Raphael. Inferring parsimonious migration histories for metastatic cancers. *Nature Genetics*, 50:718-726, 2018.

2017   M. El-Kebir, B.J. Raphael, R. Shamir, R. Sharan, S. Zaccaria, M. Zehavi and R. Zeira. Complexity and algorithms for copy-number evolution problems. *Algorithms for Molecular Biology*, 12(1), 2017. Extended version of selected papers that appeared in Workshop on Algorithms in Bioinformatics (WABI) 2016.

2016	The Computational Pan-Genomics Consortium. Computational pan-genomics: status, promises and challenges. *Briefings in Bioinformatics*, bbw089, 2016.
	M. El-Kebir*, G. Satas*, L. Oesper, B.J. Raphael. Inferring the Mutational History of a Tumor using Multi-State Perfect Phylogeny Mixtures. *Cell Systems*, 3(1):43-53, 2016. Extended version of selected papers that appeared in Research in Computational Molecular Biology (RECOMB) 2016.
	A. May, B. W. Brandt, M. El-Kebir, G. W. Klau, E. Zaura, W. Crielaard, J. Heringa and S. Abeln. metaModules identifies key functional subnetworks in microbiome-related disease. *Bioinformatics*, 32(11):1678–1685, 2016.
2015	M. El-Kebir, J. Heringa and G. W. Klau. Natalie 2.0: Sparse Global Network Alignment as a Special Case of Quadratic Assignment. *Algorithms*, 8 (4), 1035-1051, 2015.
	M. El-Kebir*, H. Soueidan*, T. Hume*, D. Beisser, M. Dittrich, T. Müller, G. Blin, J. Heringa, M. Nikolski, L. F. A. Wessels, G. W. Klau. xHeinz: An algorithm for mining cross-species network modules under a flexible conservation model. *Bioinformatics*, 31(19):3147–3155, 2015.
	M. El-Kebir*, L. Oesper*, H. Acheson-Field, B. J. Raphael. Reconstruction of clonal trees and tumor composition from multi-sample sequencing data, *Bioinformatics*, 31(12):i62-i70, 2015. Special issue for Intelligent Systems for Molecular Biology (ISMB) 2015.
2014	K. Dinkla*, M. El-Kebir*, C.-I. Bucur, M. Siderius, M. J. Smit, M. A. Westenberg, and G. W. Klau. eXamine: Exploring annotated modules in networks. *BMC Bioinformatics*, 15(1):201, 2014. Extended version of selected papers that appeared in Symposium on Biological Data Visualization (BioVis) 2014.
	M. El-Kebir*, B. W. Brandt*, J. Heringa, and G. W. Klau. NatalieQ: A web server for protein-protein interaction network querying. *BMC Systems Biology*, 8(1):40, 2014.
2013	M. El-Kebir*, T. Marschall*, I. Wohlers*, M. Patterson, J. Heringa, A. Schönhuth, and G. W. Klau. Mapping proteins in the presence of paralogs using units of coevolution. *BMC Bioinformatics*, 14(Suppl 15):S18, 2013. Special issue for RECOMB Comparative Genomics (RECOMB-CG) 2013.
	M. El-Kebir*, M. van der Kuip*, A. M. van Furth, and D. E. Kirschner. Computational modeling of tuberculous meningitis reveals an important role for tumor necrosis factor-α. *Journal of Theoretical Biology*, 328(C):43–53, Mar. 2013.
	S. Canzar*, M. El-Kebir*, R. Pool, K. Elbassioni, A. K. Malde, A. E. Mark, D. P. Geerke, L. Stougie, and G. W. Klau. Charge Group Partitioning in Biomolecular Simulation. *Journal of Computational Biology*, 20(3):188–198, Mar. 2013. Extended version of Research in Computational Molecular Biology (RECOMB) 2013 paper.
2011	S. Marino, M. El-Kebir, and D. Kirschner. A hybrid multi-compartment model of granuloma formation and T cell priming in Tuberculosis. *Journal of Theoretical Biology*, 280(1):50–62, July 2011.
	M. Fallahi-Sichani, M. El-Kebir, S. Marino, D. E. Kirschner, and J. J. Linderman. Multiscale Computational Modeling Reveals a Critical Role for TNF-Receptor 1 Dynamics in Tuberculosis Granuloma Formation. *The Journal of Immunology*, 186(6):3472–3483, 2011.

Peer-reviewed conferences

Journal papers that subsequently appeared in conference proceedings first have been omitted in the following.

2023 | **S. Ivanovic**[*] and <u>M. El-Kebir</u>[†]. *Modeling and Predicting Cancer Clonal Evolution with Reinforcement Learning.* RECOMB 2023. Annual International Conference on Research in Computational Molecular Biology, Istanbul, Turkey, April 16-19, 2023. Invited to *Genome Research.*

2022 | <u>M. El-Kebir</u>[†], Q. Morris, L. Oesper, and S.C. Sahinalp. *Emerging Topics in Cancer Evolution*, PSB 2022, Pacific Symposium on Biocomputing. The Big Island of Hawaii, January 3-7, 2022.

2019 | **S. Christensen**, M. D. M. Leiserson and <u>M. El-Kebir</u>[†]. *PhySigs: Phylogenetic Inference of Mutational Signature Dynamics.* PSB 2020, Pacific Symposium on Biocomputing, Mutational Signatures session, The Big Island of Hawaii, January 3-7, 2020.
P. Sashittal and <u>M. El-Kebir</u>[†]. *SharpTNI: Counting and Sampling Parsimonious Transmission Networks under a Weak Bottleneck.* RECOMB-CG 2019, RECOMB Comparative Genomics, Montpellier, France, October 1-4, 2019.

2017 | M. S. Engler, <u>M. El-Kebir</u>, A. E. Mark, D. P. Geerke and G. W. Klau. *Enumerating common molecular substructures.* GCB 2017, German Conference on Bioinformatics, Tübingen, Germany, September 18-20, 2017.

2011 | <u>M. El-Kebir</u>, J. Heringa and G. W. Klau. *Lagrangian relaxation applied to sparse global network alignment.* PRIB 2011, International Conference on Pattern Recognition in Bioinformatics, Delft, The Netherlands, November 2-4 2011.
S. Canzar[*] and <u>M. El-Kebir</u>[*]. *A mathematical programming approach to marker-assisted gene pyramiding..* WABI 2011, Workshop on Algorithms in Bioinformatics, Saarbrücken, Germany, September 5-7 2011, Lecture Notes in Computer Science 6833: 26–38.

2007 | S. Schlobach, E. Blaauw, <u>M. El-Kebir</u>, A. ten Teije, F. Van Harmelen, *et al. Anytime classification by ontology approximation.* NFRSW 2007, New forms of reasoning for the Semantic Web: scalable, tolerant and dynamic, pages 60–74, 2007.

Peer-reviewed workshop

2020 | <u>M. El-Kebir</u>[†], **J. Oh**, **Y. Qi** and **P. Sashittal**. *Counting and Sampling Problems in Computational Biology*. Model Counting Workshop (MCW) 2020, part of the International Conference on Theory and Applications of Satisfiability Testing (SAT) 2020, July 9th, 2020, virtual.

Other (editorials, opinions, etc.)

2022 | B. Berger, D. Tian, W.V. Li, <u>M. El-Kebir</u>, A.I. Tomescu, R. Singh, N. Beerenwinkel, Y. Li, C. Boucher and Z. Bar-Joseph. What Are the Keys to Succeeding as a Computational Biologist in Today's Research Climate? *Cell Systems 13, no. 10 (October 2022): 781–85.*

Invited talks

2022 | **Accurate Identification of Transcription Regulatory Sequences and Genes in Coronaviruses**
International Society for Computational Biology (ISCB) / Society for Molecular Biology and Evolution (SMBE) Webinar. November 2022.

Combinatorial Algorithms for Tumor Phylogenetics
CRC Integromics Group at MD Anderson Cancer Center. Houston, Texas, October 2022.

Accurate Identification of Transcription Regulatory Sequences and Genes in Coronaviruses
Computational Genomics Summer Institute (CGSI) 2022. Los Angeles, California, July 2022.

2021 **Combinatorial Algorithms for Tumor Phylogenetics**
DePaul University. Department of Computer Science. Chicago, Illinois, October 2021 (virtual).

Combinatorial Algorithms for Tumor Phylogenetics
Sorbonne Université. Laboratory of Computational and Quantitative Biology. Paris, France, May 2021 (virtual).

Combinatorial Algorithms for Tumor Phylogenetics
Florida State University. Department of Computer Science. Tallahassee, FL, February 2021 (virtual).

2020 **Characterizing SARS-CoV-2 viral diversity within and between hosts**
Talk at the 9th Workshop on Computational Advances in Molecular Epidemiology (CAME), virtual, October 2020.

2019 **Implications of Non-uniqueness of Solutions in Cancer Phylogenetics**
University of Maryland College Park. Center for Bioinformatics and Computational Biology, College Park, MD, October, 2019.

Summarizing the Solution Space in Tumor Phylogeny Inference by Multiple Consensus Trees
Talk at RECOMB Computational Cancer Biology (RECOMB-CCB) 2019, Washington, DC, May 2019.

SPhyR: Tumor Phylogeny Estimation from Single-Cell Sequencing Data under Loss and Error
Late-breaking research talk at Intelligent Systems for Molecular Biology (ISMB) 2018, Chicago, IL, July 2018.

Inferring Parsimonious Migration Histories for Metastatic Cancers
Highlight talk at Intelligent Systems for Molecular Biology (ISMB) 2018, Chicago, IL, July 2018.

Combinatorial Algorithms for Tumor Phylogenetics
Indiana University Bloomington, Bloomington, IN, April 2018.

2017 **Combinatorial Algorithms for Tumor Phylogenetics**
Carnegie Mellon University, Pittsburgh, PA, May 2017.

Combinatorial Algorithms for Tumor Phylogenetics
Rice University, Houston, TX, April 2017.

Combinatorial Algorithms for Tumor Phylogenetics
University of Illinois at Urbana-Champaign, Urbana, IL, April 2017.

2016 **Inferring the Mutational History of a Tumor using Perfect Phylogeny Mixtures**
Univ. degli Studi di Milano-Bicocca, Milan, Italy, September 2016.

An MILP formulation for the Variant Allele Frequency Factorization Problem
Simons Institute for the Theory of Computing, Integer Linear Programming in Computational Biology workshop, Berkeley, CA, May 2016.

Multi-State Perfect Phylogeny Mixture Deconvolution and Applications to Cancer Sequencing
Simons Institute for the Theory of Computing, Computational Cancer Biology workshop, Berkeley, CA, February 2016.

2014	**Solving the maximum-weight connected subgraph problem**
	Exact algorithms for bioinformatics meeting, Shonan Village Center, Japan, March 2014.
	Enumerating all maximal common connected subgraphs
	Exact algorithms for bioinformatics meeting, Shonan Village Center, Japan, March 2014.
2012	**Marker-assisted gene pyramiding**
	Rijk Zwaan B.V. (plant breeding company), Fijnaart, The Netherlands, December 2012.
	A web server for topology-aware global protein-protein interaction network comparison
	Talk at the Netherlands Bioinformatics conference (NBIC 2011), April 2012.
2011	**Lagrangian relaxation applied to network alignment**
	Talk at the International Conference on Operations Research (OR 2011), Zürich, Switzerland, September 2011.

Teaching Experience

Fall 2022	**Introduction to Bioinformatics**, CS 466 at UIUC (50 students).
Spring 2022	**Computational Cancer Genomics**, CS 598MEB at UIUC (11 students).
Fall 2021	**Introduction to Bioinformatics**, CS 466 at UIUC (55 students).
Spring 2021	**Computational Cancer Genomics**, CS 598MEB at UIUC (13 students).
Fall 2020	**Introduction to Bioinformatics**, CS 466 at UIUC (62 students).
Spring 2020	**Computational Cancer Genomics**, CS 598MEB at UIUC (16 students).
Fall 2019	**Introduction to Bioinformatics**, CS 466 at UIUC (46 students).
Spring 2019	**Computational Cancer Genomics**, CS 598MEB at UIUC (13 students).
Fall 2018	**Introduction to Bioinformatics**, CS 466 at UIUC (36 students).

PhD thesis adviser

Former:

2022	Juho Kim, Ph.D. student, Electrical and Computer Engineering, co-advised with Sanmi Koyejo, graduated 7/2022. Thesis title: Probabilistic subclonal reconstruction for cancer.
2020	Sarah Christensen, Ph.D. student, Computer Science, co-advised with Tandy Warnow, graduated 12/2020. Thesis title: Algorithms for phylogenetic tree correction in species and cancer evolution. Now Technical Program Manager at D.E. Shaw Research.

Current:

Chuanyi Zhang, Ph.D student, Electrical and Computer Engineering, co-advised with Idoia Ochoa, expected graduation: 2023. Tentative thesis title: Algorithms for germline, cancer and viral genomics.

Yuanyuan Qi, Ph.D. student, Computer Science, expected graduation: 2024. Topic: tumor phylogeny estimation from bulk DNA-seq data.

Leah Weber, Ph.D. student, Computer Science, expected graduation: 2024. Topic: single-cell cancer genomics.

Mrinmoy Roddur, Ph.D. student, Computer Science, expected graduation: 2025. Topic: metastasis and transmission reconstruction.

Stefan Ivanovic, Ph.D. student, Computer Science, expected graduation: 2026. Topic: application of machine learning to cancer phylogenetics.

Nicole Dong, Ph.D. student, Computer Science, expected graduation: 2027. Topic: cancer phylogenetics.

MS thesis adviser

Former:

2021   Palash Sashittal, MS student, Computer Science, graduated 8/2021. Thesis title: Algorithms for infection and cancer genomics. Now postdoc in the Department of Computer Science at Princeton University.

2020   Nuraini Aguse, MS student, Computer Science, graduated 5/2020. Thesis title: Methods to summarize and reduce the solution space of tumor phylogeny inference. Now Software Engineer at Google.

         Jiaqi Wu, MS student, Computer Science, graduated 5/2020. Thesis title: Deciphering the heterogeneity and spatial architecture of tumors. Now Software Engineer at Facebook.

2019   Shunping Xie, MS student, Computer Science, graduated 5/2019. Towards characterizing the solution space of the 1-Dollo phylogeny problem. Now Software Engineer at Pony.ai.

Graduate research mentor

While the previous sections listed advised thesis students, the following sections list research activities with more limited scope, typically in the form of a single-semester independent study. Any resulting publications are indicated.

Former:

2022   Xinyu Gu, MS Bioinformatics student, Computer Science. Independent study in Spring 2022 and Fall 2022. Topic: mRNA design. Now continuing MS Bioinformatics student at UIUC CS.

2021   Sophia Yang, Ph.D. student, Computer Science. Independent study in Fall 2021. Topic: adapting sequence alignment algorithms for SQL query assessment in CS education. Resulted in publication at DataEd 2022 conference. Now continuing Ph.D. student at UIUC CS.

2020   James Willson, Ph.D. student, Computer Science. Independent study in Fall 2020. Topic: 1-Dollo phylogeny. Now continuing Ph.D. student at UIUC CS.

2019   Silas Hsu, Ph.D. student, Computer Science. Independent study in Spring 2019. Topic: visualization of copy-number aberrations in cancer genomics data. Resulted in publication in PLOS CB. Now continuing Ph.D. student at UIUC CS.

2018   Amitha Sandur, MS student, Computer Science (Bioinformatics). Independent study in Fall 2018. Co-advised with Kyle Schachtschneider (Research Assist. Professor, Radiology, UIC). Resulted in publication in *Frontiers in Genetics*. Now Bioinformatics Production Engineer at Natera.

Current:

         Vikram Ramavarapu, MS Bioinformatics student, Computer Science. Independent study in Spring 2023. Topic: visualization of phylogenies and transmission networks in metastasis.

         Griffen Mustion, MS Bioinformatics student, Computer Science. Independent study in Spring 2023. Topic: motif detection in viral RNA genomes.

Undergraduate research mentorship

Former:

2022   Nitya Sunkad, BS student, Computer Science. CS STARS in Fall 2022. Topic: visualization of single-cell cancer genomics data.

Keshav Gandhi, BS student, UIC Statistics. REU program at UIUC CS in Summer 2022. Topic: phylogenetic deconvolution of PolyG data.

Mayank Hirani, BS student, Computer Science. REU program at UIUC CS in Summer 2022. Co-advised with Alicia Kraay (Assist. Professor, Department of Kinesiology and Community Health). Topic: within-host evolution in acute and chronic SARS-CoV-2 infections.

Tyler Camp, BS student (Parkland College). GearUP program in Summer 2022. Topic: mining patterns of cancer clonal evolution using CloMu.

Sophie Kish, BS student, Computer Science. Independent study in Spring 2022. Topic: relationship between two-state perfect phylogeny and distance-based phylogeny estimation.

Michael Xiang, BS student, Computer Science. Independent study in Fall 2021 and Spring 2022. Topic: cancer genomics + web-based visual analytics. Co-author on two publications in PLOS CB and MBE.

Nayonika Roy, BS student, Computer Science. CS STARS in Fall 2021 and Spring 2022. Topic: analysis of cancer genomics data.

Michelle Jun, BS student, Computer Science. CS STARS in Fall 2021 and Spring 2022. Topic: analysis of cancer genomics data.

2021   Yichi Zhang, BS student, Computer Science. Independent study in Fall 2021. Topic: web-based visual analytics. Co-author on publication in MBE.

Ayesha Kazi, BS student, Computer Science. Independent study in Fall 2021. Topic: web-based visual analytics. Co-author on publication in MBE.

Zubair Lalani, BS student, Computer Science. REU program at UIUC CS in Summer 2021. Topic: user-guided segmentation in cancer genomics. Resulted in first-author publication in PLOS CB.

2020   Smarak Pattnaik, BS student, Computer Science. REU program at UIUC CS in Summer 2020. Topic: single-cell cancer phylogenetics.

Medha Patil, BS student, Computer Science. REU program at UIUC CS in Summer 2020. Topic: workflow development for SARS-CoV-2 genome analysis.

Jackie Oh, BS student, Computer Science. Independent study in Spring 2020. Topic: sampling 1-Dollo solutions. Resulted in workshop paper at Model Counting Workshop (MCW) 2020.

Current:

Samuil Donchev, BS student, Computer Science. Independent study in Fall 2022 and Spring 2023. Topic: identify *Mesoniviridae* transcription regulatory sequences and genes.

Akul Joshi, BS student, Computer Science. Independent study in Spring 2023. Topic: visualization of phylogenies and transmission networks in metastasis.

High-school research mentorship

Former:

2021	Toluwanimi Aroyo, high-school student. NIH STEP-UP in Summer 2021. Now undergraduate at Princeton University.
	Keshav Gandhi, high-school student. researcHStart program at Cancer Center at Illinois in Summer 2021. Now undergraduate at UIC Statistics + Guaranteed Professional Program Admissions (to UIC MD program).
	Raneem Saadah, high-school student. researcHStart program at Cancer Center at Illinois in Summer 2021.
2018	Eman Zwawi, high-school student. researcHStart program at Cancer Center at Illinois (Summer 2018). Now undergraduate at UIUC MCB.

Academic Service

Conference organization	○ Area co-chair for 'General Computational Biology' track (ISMB/ECCB 2023–2021) ○ Co-organizer NCI Spring School on Algorithmic Cancer Biology, March 13-19, 2023 ○ Co-organizer of 'Cancer Evolution' workshop at PSB 2022 ○ PC co-chair (WABI 2021) ○ Poster chair (RECOMB 2021)
Grant referee	○ Reviewer for Biodata Management and Analysis (BDMA) study section at NIH (2021) ○ Panel member for National Science Foundation (2022, 2021) ○ Ad hoc reviewer for National Science Foundation (2021 [2x], 2020) ○ Ad hoc reviewer for National Science Centre, Poland (2021) ○ Ad hoc reviewer for Bordeaux University, G2P (2016)
Journal referee	○ Nature Cancer (2022) ○ GigaScience (2022) ○ Biometrics (2022) ○ eBioMedicine (2022) ○ Journal of Computational Biology (2021) ○ Bioinformatics Advances (2021) ○ Microbial Genomics (2021) ○ Cell Reports Methods (2021) ○ Bioinformatics and Biology Insights (2021) ○ Genome Biology (2022 [2x], 2021, 2020) ○ Nature Communications (2022 [3x], 2021 [2x], 2020 [2x], 2019 [2x], 2018 [2x]) ○ Systematic Biology (2021, 2020) ○ Biostatistics (2020) ○ iScience (2022 [2x], 2020) ○ Journal of Graph Algorithms and Applications (2020) ○ Genome Biology (2020) ○ Cell Systems (2020 [2x], 2019) ○ PLOS Computational Biology (2022 [1x guest editor, 2x reviewer], 2021 [1x guest editor, 3x reviewer], 2020 [1x guest editor, 2x reviewer], 2019 [4x], 2018, 2011) ○ Bioinformatics (2022, 2021 [2x], 2019 [3x], 2018 [2x], 2017) ○ Communications Biology (2018) ○ Genome Research (2018 [2x]) ○ F1000Research (2018) ○ PeerJ (2018)

	○ IEEE/ACM Transactions on Computational Biology and Bioinformatics (2022, 2018, 2017) ○ BMC Bioinformatics (2017, 2014, 2012) ○ Fundamenta Informaticae (2017) ○ PLOS One (2016) ○ EURASIP Journal on Bioinformatics and Systems Biology (2016) ○ Frontiers in Bioengineering and Biotechnology (2014) ○ Computational Biology and Chemistry (2014) ○ Scientific Reports (2014) ○ Journal of Chemical Information and Modeling (2013) ○ Journal of the ACM (2011)
Conference referee	○ Research in Computational Molecular Biology (RECOMB, 2020–2017, 2015) ○ RECOMB-Comparative Genomics (RECOMB-CG, 2019) ○ Intelligent Systems in Molecular Biology (ISMB, 2018) ○ Asia Pacific Bioinformatics Conference (APBC, 2017, 2012) ○ RECOMB-Computational Cancer Biology, (RECOMB-CCB, 2017) ○ German Conference on Bioinformatics (GCB, 2012) ○ Workshop on Algorithms in Bioinformatics (WABI, 2011) ○ Symposium on Experimental Algorithms (SEA, 2011)
Conference program committee	○ Conference on Research in Computational Molecular Biology (RECOMB, 2023–2020) ○ Pacific Symposium on Biocomputing (PSB, 2020) ○ Current Trends in Theory and Practice of Computer Science (SOFSEM, 2021, 2020) ○ RECOMB-Comparative Genomics, (RECOMB-CG, 2023–2021) ○ RECOMB-Computational Cancer Biology, (RECOMB-CCB, 2023–2022, 2020–2019) ○ Intelligent Systems in Molecular Biology (ISMB, 2023–2019) ○ Workshop on Algorithms in Biology (WABI, 2022, 2021, 2019) ○ Great Lakes Bioinformatics Conference (GLBIO, 2021, 2019) ○ International Symposium on Mathematical and Computational Oncology (ISMCO, 2021–2019) ○ ACM Conference on Bioinformatics, Computational Biology, and Health Informatics (ACM-BCB, 2022–2018) ○ Computational Intelligence methods for Bioinformatics and Biostatistics (CIBB, 2019, 2017) ○ European Conference on Computational Biology (ECCB, 2019, 2016)
University of Illinois service	○ Graduate awards committee, Computer Science, UIUC (2019–present) ○ Graduate studies committee, Computer Science, UIUC (2019–present) ○ Ad hoc subcommittee to review curriculum proposal. Grainger College of Engineering, UIUC (2022) ○ Student awards committee, Computer Science, UIUC (2018–2019) ○ Admissions committee, Carle-Illinois College of Medicine, UIUC (2019)

Software

I am committed to making software developed in my lab available on GitHub under an open source license.

2023	**CloMu** [https://github.com/elkebir-group/clomu] CloMu is a neural network based algorithm for modeling and predicting cancer clonal evolution. Implementation details: Python

2022 | **CNAViz** [`https://elkebir-group.github.io/cnaviz`]
CNAViz is an interactive webtool for user-guided segmentation of tumor DNA sequencing data.
Implementation details: React, TypeScript.

CORSID [`https://github.com/elkebir-group/corsid`]
CORSID is a computational tool to simultaneously identify TRS sites, the core sequence and gene locations given an unannotated coronavirus genome sequence.
Implementation details: Python.

2021 | **JUMPER** [`https://github.com/elkebir-group/jumper`]
JUMPER solves the Discontinuous Transciption Assembly problem to infer transcripts $\mathcal{T}$ and their abundances $\mathbf{c}$ with maximum likelihood given a *Nidovirales* alignment $\mathcal{T}$ of (paired end) reads.
Implementation details: Python

PACTION [`https://github.com/elkebir-group/paction`]
PACTION solves the Parsimonious Clone Tree Integration problem of inferring clones $\Pi \subseteq \Pi_1 \times \Pi_2$, a clone tree T and proportions U given input clone trees T_1, T_2 with clones Π_1, Π_2 and proportions U_1, U_2.
Implementation details: Python

doubletD [`https://github.com/elkebir-group/doubletD`]
Tool for detecting doublets in medium to high-coverage single-cell DNA sequencing data.
Implementation details: Python

2020 | **RECAP** [`https://github.com/elkebir-group/RECAP`]
RECAP is an iterative, heuristic algorithm for solving the Multiple Choice Consensus Tree (MCCT) problem. The input is a family of sets of patient tumor phylogenies and an integer $k > 0$. RECAP then returns (i) a single expanded tumor phylogeny for each patient, (ii) an assignment of patients into k clusters, and (iii) a consensus tree for each cluster summarizing its repeated evolutionary patterns.
Implementation details: C++

Phyolin [`https://github.com/elkebir-group/phyolin`]
Phyolin is an algorithm to test for linear evolution given single-cell DNA sequencing data of a tumor. Specifically, Uuder the assumption that the phylogeny is linear, Phyolin outputs an estimated false negative rate, an inferred linear perfect phylogeny and classification of the tree topology based on the inputted threshold.
Implementation details: C++, CPLEX

TiTUS [`https://github.com/elkebir-group/TiTUS`]
TiTUS takes as input a timed pathogen phylogeny whose leaves are labeled by hosts along with epidemiological data. It counts and uniformly samples from the set of feasible interval vertex labelings of the timed phylogeny that satisfy the direct transmission constraint while supporting a weak transmission bottleneck.
Implementation details: C++, Bash

PhyDOSE [`https://github.com/elkebir-group/PhyDOSE`]
Given a set of candidate tumor phylogenies, a frequency matrix obtained from bulk data, and a confidence level, PhyDOSE computes the minimum number of single cells needed in a follow-up single-cell sequencing (SCS) experiment to determine the true phylogeny among the set of given phylogenies with the desired confidence level.
Implementation details: C++, R

13

2019 | **PhySigs** [https://github.com/elkebir-group/PhySigs]
PhySigs solves the Tree-constrained Exposure inference problem to identify a small number of exposure shifts along the edges of a given tumor phylogeny. This problem arises in the analysis of the dynamics of mutational signature exposures in cancer, and has both therapeutic and basic science applications.
Implementation details: R

SharpTNI [https://github.com/elkebir-group/SharpTNI]
SharpTNI is a tool for counting and sampling solutions from the space of parsimonious transmission networks under a weak transmission bottleneck constraint. This problem arises in phylodynamic and phylogeographic analyses.
Implementation details: C++, Bash

MCT [https://github.com/elkebir-group/MCT]
MCT contains exact and heuristic algorithms for solving the Multiple Consensus Tree problem. This problem arises in the summarization of the solution space in tumor phylogeny inference.
Implementation details: C++, CPLEX

2018 | **PMH-S** [https://github.com/elkebir-group/PMH-S]
PMH-S is a fixed parameter tractable (FPT) algorithm for the Parsimonious Migration History problem for the case where the migration graph is restricted to a tree (i.e. $\mathcal{P} = \{S\}$). The asymptotic running time is exponential in the number m of locations.
Implementation details: C++

SPhyR [https://github.com/elkebir-group/SPhyR]
SPhyR is an algorithm for reconstructing phylogenetic trees from single-cell sequencing data. SPhyR employs the k-Dollo phylogeny model, where each single-nucleotide variant can only be gained once but lost k times.
Implementation details: C++, CPLEX

2017 | **MACHINA** [http://github.com/raphael-group/machina]
MACHINA is a computational framework for inferring migration patterns between a primary tumor and metastases using DNA sequencing data.
Implementation details: C++, Gurobi

2016 | **SPRUCE** [http://github.com/raphael-group/spruce]
Reconstruction of tumor evolutionary history of single-nucleotide variants and copy-number aberrations from multi-sample bulk sequencing data.
Implementation details: C++

2015 | **AncesTree** [http://github.com/raphael-group/AncesTree]
Reconstruction of tumor evolutionary history of single-nucleotide variants from multi-sample bulk sequencing data.
Implementation details: C++, CPLEX

2014 | **eXamine** [http://apps.cytoscape.org/apps/examine]
Cytoscape 3.x app that displays set membership as contours on top of a node-link layout.
Implementation details: Java

xHeinz [https://github.com/ls-cwi/xheinz]
Identification of conserved active modules.
Implementation details: C++, CPLEX

Heinz 2.0 [https://github.com/ls-cwi/heinz]
Identification of active modules.
Implementation details: C++, CPLEX

NatalieQ [http://ibi.vu.nl/programs/natalieq]
Web server for protein-protein interaction network querying.
Implementation details: C++, Perl

2013 **Automated topology builder** [http://compbio.chemistry.uq.edu.au/atb/]
The ATB and Repository facilitates the development of molecular force fields for Molecular
Dynamics or Monte Carlo simulations of biomolecular systems.
Implementation details: C++, Python

2011 **Natalie 2.0** [https://github.com/ls-cwi/natalie]
Method for network alignment based on Lagrangian relaxation.
Implementation details: C++

Last updated on January 19, 2023