Abdurrahman Abul-Basher | Ph.D.

Postdoctoral Research Fellow, Life Sciences Centre, The University of British Columbia
2350 Health Sciences Mall – Vancouver, BC, Canada V6T 1Z3
✉ arbasher@student.ubc.ca   •   🖥 arbasher.github.io
[Linkedin](#) | [Google Scholar](#) | [Github](#) | [Twitter](#) | Citizenship: **Canadian**

Education

The University of British Columbia   **Vancouver, BC, Canada**
PhD in Bioinformatics   *2013–2020*
Thesis title: Machine Learning Methods for Metabolic Pathway Inference from Genomic Sequence Information.

Concordia University   **Montreal, QC, Canada**
MASc in Information Systems Security   *2009–2011*
Thesis title: Mining Chat Logs to Extract Information about Authors and Topics for Crime Investigation.

King AbdulAziz University (KAU)   **Jeddah, Saudi Arabia**
BSc in Computer Science   *2004–2008*
Thesis title: University Courses Timetabling System using Genetic Algorithm.

Honors and Awards

11. Faculty of Science - BCB2 Fund ($18,200 ∼ $27,000 per year), The University of British Columbia (UBC), Canada. *2017-2020*.

10. *President's Academic Excellence Initiative PhD Award*, The University of British Columbia (UBC), Canada. *2020*.

9. Four Year Fellowships (4YF) ($18,200 per year + tuition fee), The University of British Columbia (UBC), Canada. *2013-2017*.

8. Faculty of Science - Graduate Support Initiative (GSI) Fund ($8,500 per year), The University of British Columbia (UBC), Canada. *2013-2017*.

7. Faculty of Science - Graduate Student Support Program (GSSP) ($25,000 per year), McGill University, Canada. *2013-2018*. (**Declined**).

6. Faculty of Science and Engineering - Graduate Student Support Program (GSSP) ($27,000 per year), York University, Canada. *2013-2019*. (**Declined**).

5. Power Corporation of Canada Graduate Fellowships ($5,000), Concordia University, Canada. *2009-2010*.

4. Concordia Graduate Student Support Program (GSSP) ($15,000 per year), Concordia University, Canada. *2009-2011*.

3. First Honor Graduate for graduating with high GPA from King Abdulaziz University, Saudi Arabia. *2008*.

2. Faculty of Computing and Information Technology (FCIT) distinguished award, King Abdulaziz University (KAU), Saudi Arabia. *2008*.

1. King Abdulaziz University merit award ($2,000), Saudi Arabia. *2004-2008*.

Publications

In Preparation:

P1. Abdurrahman Abul-Basher, Ryan J. McLaughlin, Julia Anstett, Aditi N. Nagaraj, and Steven Hallam. "Leveraging multiple (less-trusted) sources to improve metabolic pathway prediction." *2022.*

Paper Under Review:

R2. Abdurrahman Abul-Basher and Steven Hallam. "reMap: relabeling multi-label pathway data with groups to enhance predictive performance." *ICCABS, 2021.*

R1. Abdurrahman Abul-Basher, Aditi N. Nagaraj, Ryan J. McLaughlin, Julia Anstett, and Steven Hallam. "leADS: improved metabolic pathway inference based on active dataset subsampling." *Bioinformatics, 2021.* [Impact Factor: 6.937].

Journal Papers:

J4. Abdurrahman Abul-Basher, Ryan J. McLaughlin, and Steven Hallam. "Metabolic pathway inference using non-negative matrix factorization with community detection." *Journal of Computational Biology, 2021.* [Impact Factor: 1.479, DOI: 10.1089/cmb.2021.0258].

J3. Abdurrahman Abul-Basher and Steven Hallam. "Leveraging heterogeneous network embedding for metabolic pathway prediction." *Bioinformatics, 2020.* [Impact Factor: 6.937, DOI: 10.1093/bioinformatics/btaa906].

J2. Abdurrahman Abul-Basher, Ryan J. McLaughlin, and Steven Hallam. "Metabolic pathway inference using multi-label classification with rich pathway features." *PLOS Computational Biology, 2020.* [Impact Factor: 4.475, DOI: 10.1371/journal.pcbi.1008174].

J1. Abdurrahman Abul-Basher and Benjamin CM Fung. "Analyzing topics and authors in chat logs for crime investigation." *Knowledge and information systems, 2014.* [Impact Factor: 2.822, DOI: 10.1007/s10115-013-0617-y].

Conference/Workshop Proceedings:

C1. Abdurrahman Abul-Basher, Ryan J. McLaughlin, and Steven Hallam. "Metabolic pathway inference using non-negative matrix factorization with community detection." *ICCABS, 2020.*

Non-peer Reviewed:

N2. Abdurrahman Abul-Basher. "Modeling Metabolic Pathways as Groups (with Augmentation)." *2019.* [PDF].

N1. Abdurrahman Abul-Basher, Alex Purdy, and Inanç Birol. "Event extraction from biomedical literature." p. 1-13. *bioRxiv, 2015.* [PDF].

Theses:

T2 Abdurrahman Abul-Basher. " Machine Learning Methods for Metabolic Pathway Inference from Genomic Sequence Information." *2020*. Supervisor: Professor Steven J. Hallam. [PDF].
T1. Abdurrahman Abul-Basher. " Mining Chat Logs to Extract Information about Authors and Topics for Crime Investigation" *2011*. Supervisor: Professor Benjamin C. M. Fung. [PDF].

Research and Professional Experience

5. Postdoctoral Research Fellow   The University of British Columbia
↪ *Hallam Lab, Vancouver, BC, Canada*   *2020–Current*
Description: Developed an ensemble-based metabolic pathway prediction ("mltS") to improve both the performance and robustness of pathway predictions using concepts borrowed from theoretical machine learning and meta-learning approaches. The rational idea is to estimate a "reliability" score associated with each member in a multi-label ensemble, and then use these scores to optimize the inference of the metabolic pathways for a newly sequenced genome. This model will be used to construct environmental pathway genome databases from environmental sequence information.
Advisor: Dr. Steven J. Hallam.

4. PhD Research Student   The University of British Columbia
↪ *Hallam Lab, Vancouver, BC, Canada*   *2016–2020*
Description: Proposed multiple solutions to predicting metabolic pathways from genomic sequence information (over 5000 organismal genomes) at different levels of complexity and completion using supervised and unsupervised machine/deep learning algorithms (e.g. triUMPF and pathway2vec). My work was the first to establish the link between the pathway prediction framework with the downstream pathway enrichment analysis method using machine learning approaches. Furthermore, some of these proposed treatments are extensible to closely associated studies in bioinformatics and machine learning (e.g. learning from partially-labeled network-structured data).
Advisors: Dr. Steven J. Hallam (primary advisor), Dr. Aria S. Hahn, and Dr. Kishori M. Konwar.

3. Research Assistant   Genome Sciences Centre
↪ *Birol Lab, Vancouver, BC, Canada*   *2013–2016*
Description: Conducted large-scale data analysis from PubMed articles using cutting-edge developments in the domain of biomedical natural language processing (BioNLP). These include building a discourse processing framework using convolutional neural networks to summarize and retrieve articles related to patient genomic and mutation profiles, addressing an efficient clustering module based on Nyström algorithm with subsampling technique, and leveraging the idea of representation learning to characterize various types of relations (e.g. discourse).
Advisors: Dr. Inanç Birol (primary advisor), Dr. Victoria A. Stuart, and Dr. Djallel Bouneffouf.

2. Part-time Data Scientist   BigOui Marketing Inc.
Montreal, QC, Canada   *2012–2014*
Description: Proposed two predictive modeling techniques (topic and graph based) to connect subscribers (BigOui) with local (Montreal) merchants and to perform quantitative and qualitative assessment of BigOui.

1. Research Assistant   Concordia University
Data Mining and Security (↪ DMaS) Lab, Montreal, QC, Canada   *2009–2011*

Description: I was the major contributor in a team of 4 members to building a framework to analyze inappropriate messages in chat logs (over 1000 logs) and to discover relevant subjects using a combination of rule-based and graphical modeling techniques. I specifically designed and proposed two models (LDA-TOT and A-TOT) which can compute the contribution of each participant for each discussed subject in a chat while discovering participants' roles in a chat. Moreover, the two models have built-in functionality to characterize the evolution of subjects along with participants over time.
Advisor: Dr. Benjamin C. M. Fung.

Teaching Experience

4. DSCI 571 Supervised Learning I, Master of Data Science, The University of British Columbia (1 semester).
3. DSCI 573 Feature and Model Selection, Master of Data Science, The University of British Columbia (1 semester).
2. DSCI 575 Advanced Machine Learning, Master of Data Science, The University of British Columbia (1 semester).
1. CPIT 201 Introduction to Computing, Computer Science, King Abdulaziz University (3 semesters).

Mentoring Experience

4. Aditi Nagaraj, MSc student, The University of British Columbia (2 months).
3. Vaibhav Grover, BSc student, Indian Institute of Technology (3 months).
2. Nathan Nastili, BSc student, Simon Fraser University (4 months).
1. Alex Purdy, BSc student, University of Victoria (12 months).

Invited Talks/Presentations

6. Boston Children's Hospital, Harvard Medical School, Harvard University, 2021, Virtual Seminar, *2021*. **Title:** Metabolic pathway prediction using non-negative matrix factorization with improved precision from (meta)genomic sequence information.
5. Center for Protein Degradation, Dana-Farber Cancer Institute, Harvard Medical School, Harvard University, 2021, Virtual Seminar, *2021*. **Title:** Metabolic pathway prediction using non-negative matrix factorization with improved precision.
4. International Conference on Computational Advances in Bio and medical Sciences (ICCABS), Virtual Conference, *2020*. **Title:** Metabolic pathway inference using non-negative matrix factorization with community detection.
3. BIOF, IOP and GSAT programs (B.I.G.) retreat, Vancouver, BC, *2019*. **Title:** Leveraging heterogeneous network embedding for metabolic pathway prediction.
2. BIOF, IOP and GSAT programs (B.I.G.) retreat, Vancouver, BC, *2018*. **Title:** Metabolic pathway inference using multi-label classification with rich pathway features.
1. Centre for Microbial Diversity and Evolution (CMDE) retreat, Victoria, BC, *2016*. **Title:** Machine learning approach to recovering metabolic pathways from metagenomics sequences.

Research Projects

My research focuses on analyzing large-scale multiomics data from single/multi organismal genomes to study microbial interactions in an ecosystem. In particular, I proposed several machine learning algorithms (i.e., metabolic pathway prediction algorithms) to address the role of microbial community complexity in shaping host fitness. Previously, I developed multiple tractable solutions to analyze large-scale PubMed articles using cutting-edge artificial intelligence technologies. Below

is a summary of my research accomplishments:

6. Proposed a unified architecture (**mltS**) that integrates the results of a flexible number of pathway prediction algorithms to recover an optimum number of metabolic pathways for single/multi organismal genomes.

5. Developed multiple solutions using **novel artificial intelligence algorithms** to predict metabolic pathways from genomic sequence information (over 8000 organismal genomes) at different levels of complexity and completion using supervised and unsupervised machine/deep learning algorithms. These include: leADS, reMap, triUMPF, pathway2vec, and mlLGPR. These tools are in the process to be integrated into MetaPathways v3 which is a popular tool to generate environmental pathway genome database from multi-organismal genome sequence information.

4. Designed a discourse parsing algorithm using **convolutional neural networks** to summarize and retrieve articles related to patient genomic and mutation profiles.

3. Developed an **event extraction algorithm** to recover structured representations of biological events from PubMed articles with associated attributes and properties, characterized by complex, nested argument structures involving several entities or recursively embedded relations.

2. Collaborated in developing an efficient clustering module based on **Nyström algorithm** with sub-sampling technique (MSSS), and leveraging the idea of representation learning to characterize various types of discourse relations.

1. Proposed a framework to analyze inappropriate messages in chat logs and discover relevant subjects using a combination of **heuristic rule-based and graphical modeling techniques**. These include LDA-TOT & A-TOT that compute the contribution of each participant for each discussed topic in a chat while discovering participants' roles in a chat.

Affiliations

5. Hallam Lab, Microbiology & Immunology, The University of British Columbia, Vancouver, BC, Canada (2016–Current).

4. Graduate Student Society, The University of British Columbia, Vancouver, BC, Canada (2013 - 2020).

3. BC Cancer Canada's Michael Smith Genome Sciences Centre, Vancouver, BC, Canada (2013 - 2016).

2. The National Cyber-Forensics and Training Alliance, Concordia University, Montreal, QC, Canada (2009 - 2011).

1. Graduate Students' Association, Concordia University, Montreal, QC, Canada (2009 - 2011).

Skills

Programming Languages

- **Proficient:** Python • R • BASH • Shell script • C#
- **Exposure:** C++ • SQL • MATLAB

Packages and Toolkits

- I have more than 3 years experience working with various downstream genomics analysis tools such as: RDKit • GSEA • g:Profiler • Cytoscape • EnrichmentMap • Phred • RAST • GenSAS • MG-RAST • Pathway-tools • EggNOG • BLAST • KAAS • rapsearch 2 • MinPath • MetaBin • KEGG • MetaCyc • Pathway Commons • WikiPathways • ConsensusPathDB • Gene Ontology (GO) • Molecular Signatures Database (MSigDB) • scikit-bio
- I have more than 5 years experience working with various visualization tools and machine/deep

learning packages such as: tensorflow • scikit-learn • pandas • Plotly • Altair • Matplotlib • MALLET • Gensim • LSTM • GRU • RNN • CNN

Miscellaneous

○ I have in-depth experience with pathway reconstruction problem from genomics data while also exploring solutions to extend pathway prediction and analysis from proteomics and metabolomics data

○ I also have a keen interest to address concerns about multi-label classification. These include: label correlations • representation learning of labels • noise in the instance/label spaces • inductive bias • active dataset subsampling • knowledge distillation • imbalanced and tail labels • partial multi-label problem • a combination of multi-instance and multi-view problem • bagging labels (grouping labels into non-disjoint sets) • high inference cost for the extreme multi-label case

References

References available upon request.