

# Abdur Rahman M. A. Basher

Vancouver, BC – Canada

✉ [abdurrahman.mohd@gmail.com](mailto:abdurrahman.mohd@gmail.com) • [arbasher.github.io](https://github.com/arbasher)

Preferred name: Abdurrahman Abul-Basher • Citizenship: Canadian

## Research Interests

---

My research focuses on designing mathematical models to address challenges associated with microorganisms and their interactions (e.g. metabolic pathways) in an ecosystem. Towards achieving this goal, I adapt and extend solutions from multi-label learning to infer biochemical reactions and metabolic pathways for single and multi-organismal genomes. Many of the proposed methodologies were already seen to achieving the state-of-the-art results in both metabolic pathway prediction and multi-label learning. Furthermore, I am interested in building a community-driven integrated ecosystem encompassing genomics, proteomics, and metabolomics data to solve the pathway reconstruction problem or discovering novel pathways. This has potential benefits for many downstream bioinformatics applications (e.g. differential pathway analysis, understanding disease mechanisms and searching for new therapeutic targets, elucidating the metabolic functions of an organism, and understanding microorganisms interactions and their impacts on the host, and vice versa).

## 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.

## Research Experience

---

**Research associate**

**University of British Columbia**

*Bioinformatics (🏠 [Hallam](#)) Lab, Vancouver, BC, Canada*

2016–2020

Research: Proposed multiple solutions to predicting metabolic pathways from genomic sequence information at different levels of complexity and completion using supervised and semi-supervised machine learning algorithms.

Advisors: Dr. Steven J. Hallam (primary advisor), Dr. Aria S. Hahn, and Dr. Kishori M. Konwar.

**Research intern**

**Genome Sciences Centre**

*Bioinformatics Technology (🏠 [Birol](#)) Lab, Vancouver, BC, Canada*

2013–2016

Research: 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 Nystrom 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.

### **Part-time data scientist**

*Montreal, QC, Canada*

### **BigOui Marketing Inc.**

*2012–2014*

Role: 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.

### **Research associate**

*Data Mining and Security (DMaS) Lab, Montreal, QC, Canada*

### **Concordia University**

*2009–2011*

Research: Proposed algorithms based on topic models to characterize topics, compute the contribution of authors in these topics, and study the transitions of these topics over time.

Advisors: Dr. Benjamin C. M. Fung (primary advisor).

## **Publications**

---

### **Accepted/Under Review**.....

6: Abdur Rahman M. A. Basher and Steven Hallam. “Multi-label Pathway Prediction based on Active Dataset Subsampling.” *Bioinformatics*, 2020 [under review].

5: Abdur Rahman M. A. Basher and Steven Hallam. “reMap: Relabeling Metabolic Pathway Dataset with Bags to Enhance Predictive Performance.” *Bioinformatics*, 2020 [under review].

4: Abdur Rahman M. A. Basher, Ryan J. McLaughlin, and Steven Hallam. “Metabolic Pathway Inference using Non-negative Matrix Factorization with Community Detection.” *Bioinformatics*, 2020 [under review].

3: Abdur Rahman M. A. Basher and Steven Hallam. “Leveraging Heterogeneous Network Embedding for Metabolic Pathway Prediction.” *Bioinformatics*, 2020.

2: Abdur Rahman M. A. Basher, Ryan J. McLaughlin, and Steven Hallam. “Metabolic Pathway Inference using Multi-Label Classification with Rich Pathway Features.” *PLOS Computational Biology*, 2020.

1: Abdur Rahman M. A. Basher and Benjamin CM Fung. “Analyzing Topics and Authors in Chat Logs for Crime Investigation.” *Knowledge and information systems*, 2014.

### **In Preparation/Non-peer Reviewed**.....

4: Abdur Rahman M. A. Basher. “mltS v2 [TO BE ADDED].” 2020.

3: Abdur Rahman M. A. Basher and Steven Hallam. “mltS [TO BE ADDED].” 2020.

2: Abdur Rahman M. A. Basher and Steven Hallam. “Modeling Metabolic Pathways as Bags (with Augmentation).” 2020.

1: Abdur Rahman M. A. Basher, Alex Purdy, and Inanç Birol. (2015). "Event Extraction from Biomedical Literature."  *biorxiv*. 1-13.

## Presentations

---

3: Abdur Rahman M. A. Basher and Steven Hallam. "Leveraging Heterogeneous Network Embedding for Metabolic Pathway Prediction.", BIOF, IOP and GSAT programs (B.I.G.) retreat, Vancouver, BC, 2019.

2: Abdur Rahman M. A. Basher and Steven Hallam. "Metabolic Pathway Inference using Multi-Label Classification with Rich Pathway Features.", BIOF, IOP and GSAT programs (B.I.G.) retreat, Vancouver, BC, 2018.

1: Abdur Rahman M. A. Basher, Connor Morgan-Lang, and Steven Hallam. "Machine Learning Approach to Recovering Metabolic Pathways from Metagenomics Sequences", Centre for Microbial Diversity and Evolution (CMDE) retreat, Victoria, BC, 2016.

## Fellowships and Awards

---

10: Faculty of Science - BCB2 Fund (\$18,200 ~ \$27,000 per year), The University of British Columbia (UBC), Canada. 2017-2020.

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.

## Teaching Experience

---

### Teaching Assistant.....

The University of British Columbia

Vancouver, BC, Canada

2016–2017

I was a TA for the following Master of Data Science ([MDS](#)) courses:


- DSCI 571 Supervised Learning I
- DSCI 573 Feature and Model Selection
- DSCI 575 Advanced Machine Learning

## Affiliations

---

Graduate Student Society


University of British Columbia

 [GSS](#), Vancouver, BC, Canada

2013–2020

Canada's Michael Smith Genome Sciences Centre

BC Cancer Agency

 [GSC](#), Vancouver, BC, Canada

2013–2016

## Skills

---

### Programming Languages.....

Python • R • MATLAB •  $\text{\LaTeX}$  • C++ • C#

### Packages and Toolkits.....

- 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 • GMTK • Stanford CoreNLP • BANNER • Gensim • NLTK • LSTM (GRU) • RNN • CNN • scikit-bio
- I have more than 3 years experience working with various downstream genomics analysis tools such as: Phred • RAST • GenSAS • MG-RAST • Pathway-tools • EggNOG • BLAST • KAAS • rapsearch 2 • MinPath • MetaBin

### Miscellaneous.....

- I proposed or currently addressing 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
- I have in-depth experience with pathway reconstruction problem from genomics data while also exploring solutions to extend pathway analysis from proteomics and metabolomics data

## References

---

Available upon request.