

Chantriolnt-Andreas Kapourani

Institute of Genetics and Cancer
Edinburgh, EH4 2XU, UK
✉ C.A.Kapourani@ed.ac.uk
✉ andreaskapou.github.io

The aim of my career is to pursue cross-disciplinary research in the intersection of data science and biomedicine

Research and Professional Experience

2018–Current **Cross-disciplinary fellow (XDF)**, MRC Human Genetics Unit and School of Informatics, University of Edinburgh.

Advisors Dr Catalina Vallejos and Prof Neil Henderson

Research project: *Machine learning and statistical modelling of single-cell genomics*

The aim of my post-doctoral project is to develop robust machine learning methods to understand the relationship between tissue localization and gene expression variation, as well as to study how location and function interact to regulate the transition from healthy to pathological disease states. We use multi-modal single cell technologies to decode the molecular mechanisms regulating liver fibrosis and regeneration. These methods will be immediately applied to determine the cellular composition and molecular mechanisms associated to cells co-located within areas of scarring during human liver injury – a condition that affects more than 844 million people worldwide.

2015–2021 **Teaching Assistant / Student Supervisor**, School of Informatics, University of Edinburgh.

- Supervisor: During my XDF fellowship I co-supervised (with Prof Sanguinetti) an MInf student for a 2-year research project on 'Cancer Data Science and Epigenetics'. During my PhD I was involved in the supervision of two Erasmus students working on machine learning for computational biology.
- Tutor: Introductory Applied Machine Learning.
- Demonstrator: Informatics 2B - Algorithms, Data Structures & Learning. Created course lab contents: <http://www.inf.ed.ac.uk/teaching/courses/inf2b/learnLabSchedule.html>
- Marker: Bioinformatics 1 and Informatics 2B.

2013–2014 **Junior Developer**, DaXtra Technologies.

DaXtra Technologies is a specialist company in high accuracy CV and job parsing technologies. I was responsible for adapting the company's software to client needs such as: integration with 3rd party Database Management Systems, integration with web services and online job boards including Indeed and LinkedIn.

Computational Genomics Software

- SeuratPipe SeuratPipe streamlines and enables reproducible computational analysis of single-cell genomics data by wrapping common tasks in easy to run pipelines. Currently supports sc/snRNA-seq and Visium data analyses (<https://github.com/andreaskapou/SeuratPipe>).
- scMET scMET is a hierarchical Bayesian model that shares information across cells and genomic features to robustly quantify genuine biological single-cell epigenetic heterogeneity. scMET can identify highly variable features that drive epigenetic heterogeneity, and perform differential methylation and variability analyses (<https://bioconductor.org/packages/scMET/>).
- Melissa Melissa is a Bayesian probabilistic model for jointly clustering and imputing single-cell methylation data. This is done by taking into account local correlations via a generalised linear model and global similarities using a mixture modelling approach (<https://bioconductor.org/packages/Melissa/>).
- BPRMeth BPRMeth is a probabilistic method to extract higher order features of methylation profiles, in a way that would make it easier to formally use such profiles in downstream modelling efforts, such as predicting gene expression or clustering genomic regions or cells according to their methylation profiles. (<https://bioconductor.org/packages/BPRMeth/>).

Education

- 2015–2018 **PhD in Data Science**, *School of Informatics*, University of Edinburgh.
Under the supervision of Prof Guido Sanguinetti
Thesis: *Spatial statistical modelling of epigenomic variability.*
High throughput sequencing platforms has enabled the profiling of different epigenetic marks on a genome-wide scale; however, bespoke computational methods are required to interpret these high-dimensional data and investigate the coupling between the epigenome and transcriptome. During my PhD, I developed statistical machine learning methods to model (single-cell) epigenetic data and capture spatial correlations of epigenetic marks that would uncover the interplay between genetic and epigenetic mechanisms in transcriptional regulation.
- 2014–2015 **MSc by Research in Data Science**, *School of Informatics*, University of Edinburgh.
Distinction
Dissertation: *Mixture modelling of high throughput biomedical data.*
Developed a machine learning approach for extracting higher order methylation features which were predictive of gene expression levels. Also, performed Bayesian integrative clustering of heterogeneous sources of high-throughput sequencing technologies. Due to their statistical power and flexibility, these tools could become widespread in the high throughput bioinformatics workbench for performing biomedical data analysis.
- 2012–2013 **MSc in Artificial Intelligence**, *School of Informatics*, University of Edinburgh.
Distinction
Dissertation: *Unsupervised motif discovery from acoustic time series data.*
Implemented computational strategy to discover and extract recurring patterns (i.e. linguistic phrases or words) from input speech data in an unsupervised setting.
- 2008–2012 **Ptychio in Informatics**, *Dept of Informatics and Telematics*, Harokopio University of Athens.
Grade 9.1/10 – Distinction
Dissertation: *Integration of external applications in the Unity academic social network environment.*
Extended the functionality of the Unity academic social network through integration of external applications in the form of Web Services, enabling interoperability with other information systems, using Google's OpenSocial API, a reference for developing social networking environments.
- 2010–2011 **Erasmus Student**, *Department of Computer Science*, Roskilde University, Denmark.
Project: Developed *Find Friend Location*, a distributed real-time location tracking system for Android mobiles provided by SOAP Web Services.

Scholarships and Awards

- 2021 Nominated as Supervisor of the Year for the EUSA Teaching Awards at the University of Edinburgh.
- 2016 Best Paper Award in the 15th European Conference on Computational Biology (ECCB).
- 2013 Winner of Travel App Prize (Skyscanner) University of Edinburgh (<http://easyscn.appspot.com/>). Natural Language Processing system for converting natural text to flight queries.
- 2012 Bodossaki Foundation scholarship for postgraduate studies.
- 2009–2012 Scholarship from Greek State Scholarships Foundation (IKY) for excellent performance (Highest Ranked student), for the academic years 2011-2012 (4th year), 2010-2011 (3rd year), 2009-2010 (2nd year), 2008-2009 (1st year).

List of Publications

- **Kapourani, C.A.**, Argelaguet, R., Sanguinetti, G., and Vallejos, C. A. (2021). scMET: Bayesian modelling of DNA methylation heterogeneity at single-cell resolution. *Genome Biology*, 22(1), pp.1-21.
- Gangoso, E., Southgate B., Bradley L., Rus, S., Galvez-Cancino, F., McGivern, N., **Kapourani, C.A.**, Byron, A., Ferguson, K., Morrison, G., Alfazema, N., Grant, V., Blin, C., Guc, E., Sou, I., Marques, M.A., Conde, L., Herrero, J., Brennan, P., Brandner, S., Bertone, P., Pollard, J.W., Quezada, S., Sproul, D., Margaret Frame, M., Serrels, A. and Pollard, S. (2021). Glioblastoma stem cells hijack myeloid-affiliated transcription factors via epigenetic immunoediting to elicit immune-evasion, *Cell*, 184(9), pp.2454-2470.
- Argelaguet, R., Clark, S.J., Mohammed, H., Stapel, L.C., Krueger, C., **Kapourani, C.A.**, Imaz-Rosshandler, I., Lohoff, T., Xiang, Y., Hanna, C.W., Smallwood, S., Ibarra-Soria, X., Buettner, F., Sanguinetti, G., Xie, W., Krueger, F., Gottgens, B., Rugg-Gunn, P.J., Kelsey, G., Dean, W., Nichols, J., Stegle, O., Marioni J.C. and Reik, W. (2019). Multi-omics profiling of mouse gastrulation at single-cell resolution. *Nature*, 576(7787), 487-491.
- **Kapourani, C.A.** and Sanguinetti, G. (2019). Melissa: Bayesian clustering and imputation of single-cell methylomes. *Genome Biology*, 20(1), 1-15.
- Clark, S.J., Argelaguet, R., **Kapourani, C.A.**, Stubbs, T.M., Lee, H.J., Alda-Catalinas, C., Krueger, F., Sanguinetti, G., Kelsey, G., Marioni, J.C., Stegle, O. and Reik, W., (2018). scNMT-seq enables joint profiling of chromatin accessibility DNA methylation and transcription in single cells. *Nature Communications*, 9(1), p.781.
- **Kapourani, C.A.** and Sanguinetti, G. (2018). BPRMeth: a flexible Bioconductor package for modelling methylation profiles. *Bioinformatics*.
- **Kapourani, C.A.** and Sanguinetti, G. (2016). Higher order methylation features for clustering and prediction in epigenomic studies. *Bioinformatics*, 32 (17), i405-i412. (**Best Paper Award in ECCB 2016**)
- Hatzi, O., Meletakis, G., Katsivelis, P., **Kapourani, C.A.**, Nikolaidou, M. and Anagnostopoulos, D. (2014). Extending the Social Network Interaction Model to Facilitate Collaboration through Service Provision. In *Enterprise, Business-Process and Information Systems Modeling* (pp. 94-108). Springer Berlin Heidelberg.

Conference Presentations

- Nov 2020 Oral presentation in *Biology of Genomes* meeting, Cold Spring Harbor Laboratory, New York, USA. Title: scMET: Bayesian modelling of DNA methylation heterogeneity at single-cell resolution.
- May 2020 Poster presentation in *Single cell biology* conference, Wellcome Genome Campus, Cambridge, UK. Title: scMET: Bayesian modelling of DNA methylation heterogeneity at single-cell resolution.
- May 2019 Oral presentation in *Centre for Genomic & Experimental Medicine* meeting, IGMM, Edinburgh, UK. Title: Single-cell RNA-seq for identifying cell sub-population specific cis-eQTLs.
- March 2018 Oral presentation in *Mathematical and Statistical Aspects of Molecular Biology (MASAMB)* workshop, St Andrews, UK. Title: Melissa: Bayesian clustering and imputation of single-cell methylomes. DOI: [10.7490/f1000research.1115317.1](https://doi.org/10.7490/f1000research.1115317.1)
- July 2017 Poster presentations in *International Society for Computational Biology (ISMB)* conference, Prague, Czech Republic. Title: Bayesian hierarchical modelling of single-cell methylation profiles
- May 2017 Poster presentations in *Statistical Challenges in Single-Cell Biology* meeting, Ascona, Switzerland. Title: Bayesian hierarchical modelling of single-cell methylation profiles
- Sept 2016 Oral presentation in *15th European Conference on Computational Biology (ECCB)*, Hague, Netherlands. Title: Higher order methylation features for clustering and prediction in epigenomic studies. DOI: [10.7490/f1000research.1113102.1](https://doi.org/10.7490/f1000research.1113102.1)
- Sept 2016 Poster presentation in *10th International Workshop on Machine Learning in Systems Biology (MLSB)*, Hague, Netherlands. Title: Bayesian integrative clustering of heterogeneous types of high-throughput sequencing data.
- Aug 2015 Oral presentation in *Statistical Modeling of Epigenomics and Gene Regulation (SMEGR)* workshop, Harvard University, USA. Title: Mixture modelling of DNA methylation profiles.

Peer Review

Reviewer for peer-reviewed journals including Genome Biology, Bioinformatics, PLOS Computational Biology, Statistical Applications in Genetics and Molecular Biology and IEEE/ACM Transactions on Computational Biology.

Public Engagement

- Ongoing Writing tutorials with accompanying code on statistical machine learning and Bayesian inference: <https://andreaskapou.github.io/blog/>.
- 2020 Involved in the **IGMM statistical seminar series**, teaching biomedical researchers how to use statistical and machine learning methods to answer research questions.
- 2018 Contributed to the **XD Challenge** public event organised by the IGMM for undergraduate and postgraduate students across the College of Science and Engineering. We provided students with challenging problems that biomedical researchers at the IGMM face, and how they could apply their computational skills to tackle them.

Data Science Skills

- ML / Stats Generalized Linear Models, Bayesian inference, Variational Inference, Deep Learning, Gaussian Processes, Approximate inference, Optimization, Latent Variable Models, Graphical Models, Mixture models, Neural Networks, Markov Chain Monte Carlo, Kernel methods, Hidden Markov Models.
- Languages R, Python, Java, C/C++, MATLAB, SQL, HTML, JavaScript.

Languages

- Greek Native Language
- Albanian Native Language
- English International English Language Testing System (IELTS)

References

Available on request