Integration techniques for modern bioinformatics workflows
Kanterakis, Alexandros

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2018

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
Acknowledgments

Having completed this work, I would like to acknowledge a large group of people without whom, this thesis wouldn’t be a reality.

Most of all, I am grateful to my supervisor, Prof. Morris A. Swertz. I was already familiar with his groundbreaking work when, in October 2009, I found that he was looking for new PhD students. I applied in a heartbeat, knowing that this was a unique opportunity to get involved in state-of-the-art research in bioinformatics. During the years that followed, I realized that Morris is something more than a leading researcher in this area. He has a unique sense of the computational needs of the genetics and biomedical community and he provides tangible solutions that do make a difference. He taught me numerous principles in bioinformatics, which can be summarized as “genetics done right”. I am really happy to realize that today these principles are taken for granted in the field of bioinformatics, but it was people like Morris who made this happen by strongly advocating them and putting them into practice. Through Morris’ supervision, I also learned that as IT experts, it is our duty to deliver solutions to geneticists, biologists and clinicians that bring the realms of recent advances in Computer Science to their attention and make it possible to employ them. I am also thankful to Morris for personally making my moving to the Netherlands as smooth as possible and for advising me on every small practical matter. I also appreciate the fact that he offered employment to my wife Despoina. Finally, I should mention that in Morris I found a friend and a fellow computer geek, I always enjoy his company with our vivid discussions about programming, Internet culture and technology. I am hopeful that this PhD work is just the start of many more fruitful collaborations in upcoming research projects.

I am also grateful to my co-supervisor Prof. Cisca Wijmenga. Cisca has been a steady source of guidance during all these years, always pointing me in the right direction. Being a computer enthusiast, I would often underrate the significance of the biological interpretation of my work. Cisca was always there to remind me that algorithms and programs are just tools to help us reach a higher goal: to contribute to the available knowledge regarding human genetics and consequently to improve treatments for certain diseases. These considerations brought additional inspiration and encouragement to my work as there was always a strong humanitarian dimension.
What I also much admire about Cisca is her ability to supervise tens of PhD students efficiently, resolving the daily minor and major issues that every PhD student has. I, and many of my colleagues, would often feel we were losing confidence in our work, but ten minutes with Cisca was enough to re-empower ourselves, focus on “the interesting”, and regain confidence. Finally, Cisca has the very rare ability to grasp the complex and novel concepts that would appear in new research articles, breaking them down into separate parts and making them perfectly clear to me so that I could go off and use or expand them. My favorite quote from Cisca is “ask the right questions”, which is now the first thing I do when I face a new challenge.

Morris managed in a short period of time to assemble a very competent group of young bioinformaticians who were inspired by his vision and did their best to transform them into reality. This is the Genomics Coordination Center (GCC) group, which has been a very hospitable environment and I consider the experience that I gained whilst working there as my greatest professional asset. I am thankful to the following people from this group: Freerk van Dijk, from whom I learned a lot during his amazing work in creating pipelines that assembled all the terabytes of raw data from the Genome of the Netherlands (GoNL) project. Freerk has also been really good at introducing me to Dutch culture through our shared passion of rock music. Patrick Deelen was a great inspiration by being a rare blend of skillful programmer, big data analyst and biologist. George Byelas was the cluster and grid expert who showed me at first hand the challenges of High Performance Computing. Pieter Neerinckx was the Linux guru who helped me expand my knowledge on Linux programming and also on how to create so many scripts that would work... ‘automagically’. And further, Robert Wagner, my MOLGENIS first responder and also a dear friend who now frequently visits Crete and never misses the opportunity to explore some of the most distant beaches; Joeri van der Velde for showcasing so many wonderful examples of multi-omics integration; Joël Kuijer for his valuable help on the PyPedia study; and Martijn Dijkstra and Lennart Karssen for being companions on pipeline authoring. I would also like to thank the members of GCC who developed and maintained the MOLGENIS platform: Dennis Hendriksen, Erwin Winder†, Roan Kanninga, Chao Pang, Despoina Antonakaki, Marcel Kempenaar, Bart Charbon, Erik Roos and Joris Lops.

Throughout these years I was able to collaborate with a large number of people from the Genetics Department of the University Medical Center Groningen (UMCG). This gave me the opportunity to work with an extremely skillful set of biologists and geneticists and allowed me to gain a deep understanding of the challenges and needs of these fields. From this group I am thankful to: Isis Ricaño-Ponce, Javier Gutierrez-Achury, Agata Szperl, Rodrigo Almeida, Gosia Trynka, Joanna Smolonska, Barbara Hrdličková and Jihane Romanos. Apart from our fruitful collaborations, I will
always cherish the numerous moments that we shared when going out for a beer.

From the Genetics Department of UMCG I am thankful to my fellow PhD students, Harm-Jan Westra, Juha Karjalainen, Marc-Jan Bonder, Daria Zhernakova and Danny Arends. Thank you for instructing me on so many areas like genetics, programming, data analysis and statistics. Thanks also to the professors, principal investigators and post-docs: Lude Franke, Alexandra Zhernakova, Sebo Withoff, Ellen Nollen and Ritsert Jansen. Your work of integrating multiple sources of -omics data to uncover novel aspects of genetic regulation has been exemplary to me.

The Genome of the Netherlands has been by far the most endearing scientific project that I had the privilege of being a member of. This project brought together many talented researchers and created a huge and long lasting impact. I would like to thank: Prof. Paul de Bakker mainly for being so good at explaining complex (at least to me) concepts in genetic analysis; Elisa van Leeuwen for helping me create so many data analysis pipelines that showcased the value of GoNL, and Laurent Francioli for demonstrating that creating ten slides of novel findings per week is possible. Seriously Laurent, how did you do it? From the GoNL consortium, I would also like to thank Androniki Menelaou, Jessica van Setten, Carolina Medina-Gomez, Jouke Jan Hottenga, Karol Estrada and David van Enckevort.

Much of the work for this thesis took place in the Foundation for Research and Technology (FORTH) in Heraklion, Greece, whilst I was working as a collaborating researcher in the Institute of Computer Science (ICS) under the supervision of Dr. George Potamias. Dr. Potamias was kind enough to give me all the resources and time required to complete my thesis while showing a strong interest in my work in the UMCG. Along with Dr. Potamias, I would like to thank his team and my close collaborators over the last few years, Lefteris Koumakis, Kleanthi Lakiotaki and Eugenia Kartsaki. I have been lucky to have similar open and collaborative behavior shown to me by Prof. George Patrinos from the University of Patras, Greece, and by Theodora Katsila.

I am also grateful to Jackie Senior who took the time to edit the complete manuscript of this thesis. She made hundreds of corrections and even suggested factual edits that augmented not only the readability but also the scientific merit of this work.

To the members of my thesis reading committee, Prof. E.A. Valentijn, Prof. J. Heringa and Prof. H. Snieder, thank you for taking the time to read the manuscript and providing numerous suggestions and comments for improvement.

I would also like to thank Alexandra Elbakyan, the creator of the research-paper sharing platform Sci-Hub. A significant part of this work took place in a crisis-stricken country, which also affected my access to scientific journals. I would not have had access to many of the papers cited in this work without Sci-Hub.

In Groningen I had the privilege of making long-lasting friendships with many people.
So to Kostas, Georgia, Alexandros, Yannis, Angeliki, Andres, Nikos, Mayerlin, Manu, Maja, Astrid, Mariana, Athina, George, Eirini, Katerina and Manolis, thank you all for expanding my horizons in so many ways. Special thanks go to Despina Serlidaki and Kostas Tassis for hosting me for three months.

I would also like to thank Dimitris Gakis and Thaleia Konstantinou for the lovely weekends that we spent in Amsterdam and Rotterdam talking about anything from soccer to... abstract philosophy.

To my parents Nikos and Eleni, and my sister Vicky, I would like to express my gratitude for being a solid spiritual compass throughout my life. To my mother-in-law Stavroula and my mother Eleni, thank you for taking so much of your time (and patience) to take care of my children while I was working on this thesis.

Finally, this work would not have been possible without the devotion, support and encouragement of my wife Despoina. Despoina, you were always there to enlighten me, dissolve my doubts, help me with LaTeX, and even scold me when you felt that my interest was drifting away. You did that while working on your own PhD and being a caring mother. I cannot thank you enough. But above all, thank you for having our two overly cute, chaotic, bright little stars, Eleni and Demetra.