



SPOTLIGHT ON CAREERS

Interview by Joshua McGrath



**Dr. Anna
Dvorkin-Gheva**

DR. ANNA DVORKIN-GHEVA is a resident bioinformatician working within the McMaster Immunology Research Centre (MIRC) at McMaster University in Hamilton, Ontario. She completed her B.Sc. in Life Sciences at Tel Aviv University, Israel, and subsequently earned her Ph.D. in Neuroscience and Computational Biology under the supervision of Dr. Henry Szechtman at McMaster University and Dr. Ilan Golani at Tel Aviv University. Specifically, her doctoral thesis focused on the investigation of exploratory behaviour patterns in rodent models of obsessive-compulsive disorder (OCD). After completing her degree, she pursued a post-doctoral fellowship with Dr. John Hassell at McMaster University, where she leveraged computational and analytical skills to test novel therapeutics for breast cancer. Ultimately, it was during this time that an unmet demand for bioinformatic analysis across many laboratories within the McMaster health sciences community was identified. As a consequence, a position, which Anna holds today, was created. This position involves collaboration with many research groups both within and outside of MIRC to investigate a wide array of biological diseases and processes using computational approaches.

1. Can you describe your position and typical work day/week?

Broadly, my position involves designing and performing bioinformatic analyses using biological datasets in collaboration with various research groups. Some of these datasets are generated by these research groups, while others are obtained from repositories with either open or restricted access. On any given day, I may spend time answering questions pertaining to the biology of infectious diseases, cancer, allergies, or autoimmunity, among other areas. Within these subfields, I may be mining previously generated datasets, looking for patterns that identify novel and interesting areas for investigation, or alternatively, asking highly-defined research questions. In either instance, I may be utilizing one or more computational approaches to perform the analysis. For these reasons, I would say that my work days and weeks are often variable, in that the topics I study, the goals I pursue, and the tools with which I pursue them are diverse.

2. What types of analyses do you typically perform in your role?

My work spans across several fields, mainly including transcriptomics, genomics, and proteomics. Transcriptomic data obtained from microarrays, Nanostring, next-generation RNA sequencing (RNAseq), and single-cell RNAseq demonstrate gene expression patterns, which are generally used for providing a deeper understanding of biological processes behind conditions of interest, for searching potential biomarkers, or for generating hypotheses. Genomic data I work with are usually obtained from another type of next-generation sequencing – DNAseq. I use these types of data to examine the variability of specific genomic regions, including searching for mutations, polymorphisms, and sequence motifs, among other things. These genomic data can, for instance, be used to study the diversity and lineages of immunoglobulin responses. The projects involving proteomics range from analyses of data obtained from mass spectrometry to analyses of protein structures involved in protein-protein docking. Within each of these fields, I am involved with experimental design, including components such as sample size calculation, as well as processing data and performing numerous types of statistical analyses. Another aspect that I am quite enthusiastic about is data visualization, which is a crucial step in the examination and presentation of experimental results. Appropriate visualization techniques allow the reader to see results clearly and the researcher to find additional directions and questions to ask based on the data. To this end, I specialize in developing unique methods of data visualization. Ultimately, while many analyses use a core set of tools and approaches as described above, datasets frequently present unique challenges, often requiring the development of innovative approaches.

3. What is your most/least favourite part of your job?

I would say that my favourite part of the job is the diversity of both the groups I get to work with and questions I get to answer. My field of study changes on a frequent basis, often allowing me to refresh my interests and take a fresh perspective on problems. Another aspect of the job that I really enjoy is exploring large datasets, looking for new variables and patterns without a hypothesis in mind. While this type of analysis is risky, as it may or may not produce fruitful results, I find the process quite fun. I think my least favourite part of my role is the variability in workload distribution. Sometimes there can be a lot on my plate at once, while other times things can be quite slow. However, I tend to view time constraints as a challenge and am able to gain a lot of satisfaction from working under pressure. On the other hand, slow periods provide the benefit of extra time to read and learn, thus allowing me to broaden my repertoire of computational analyses.

4. How common is your role in academia/industry? Do you see the need for this type of role growing in the future?

Biological scientists, whether in academia or industry, are becoming increasingly aware of the utility of bioinformatics. In academia, many laboratories are becoming interested in pursuing these types of analysis. Often times, however, it can be inefficient to train individuals to do so, given the time and effort required to learn programming languages and specific analytical methods. The benefit that dedicated bioinformaticians provide is that they can act as a core facility and use a consolidated set of skills to provide computational support to a number of groups. This in turn allows these groups to focus their efforts within their respective fields. Ultimately, I think that the efficiency of this type of core facility-like setup will facilitate the expansion of bioinformatician roles in the future. In industry, a role that is roughly equivalent to mine would be that of a data analyst. These roles are already quite common and will likely proliferate further in the future for the reasons above.

5. Do you have any advice for graduate students who would like to pursue a career in bioinformatics?

First of all, I would say that all graduate students in the health sciences should try to develop a fundamental understanding of the computational techniques that are used in their respective fields. Although knowing the specific methodologies of these techniques may not be necessary for everyone, having a basic understanding about which techniques are used, when they should be used, why they are used, and the broad strokes of how they work will undoubtedly help in critiquing experimental rationales and protocols. For students interested in pursuing bioinformatics-based work specifically, I would recommend to begin by learning to code using languages such as R, Python, or Matlab. This is essential, as bioinformatic analyses are conducted through programming. Further to that, I would encourage students to gain literacy in more than one coding language, as each has their advantages for specific applications and you cannot always predict what language other researchers code their programs in. Finally, students should identify and take courses that are designed to teach the techniques and analyses that they are interested in.