

5000-dimensional vectors: analyzing large data sets in biochemistry

Dr. Stephen Rader¹

Living cells respond to stress and changing environmental conditions by turning genes on and off. For example, when nitrogen becomes limiting for growth, an organism might increase the expression of proteins that scavenge nitrogen from the environment. Due to breakthroughs in sequencing technology and computational power, it is now routine to monitor the expression of all of the genes in a cell simultaneously by measuring the number of RNA copies of each that are made under each different condition. Analysis then requires comparing gene expression under each condition and looking for differences. This requires a rigorous statistical model to normalize variabilities in data collection and allow for meaningful comparison. I will present some recent examples from my lab in which we exposed a unicellular red alga to nutrient starvation, varying temperature, toxic metals, etc, and explain possible applications of this kind of information.

¹UNBC Biochemistry and Molecular Biology, University of Northern British Columbia, Prince George, B.C., V2N4Z9, Canada (rader@unbc.ca).