



Sorin Istrail, Ph.D.

Julie Nguyen Brown Professor of Computational
and Mathematical Sciences

Professor of Computer Science

Department of Computer Science

Editor-in-Chief, *Journal of Computational Biology*

March 26, 2012

BJ Morrison McKay, Executive Officer
International Society for Computational Biology

Dear BJ Morrison McKay,

It is with great enthusiasm and honor that I write this letter supporting the proposal that Professor Michael Waterman be awarded the National Medal of Science.

I also write with humility, for how can I do justice in discussing the achievements of the father of the field, an influential researcher, and a mentor to so many in computational biology, a.k.a. bioinformatics? To speak of Professor Waterman is to speak of what matters most in this new and exciting field. He stands on high ground with an unmatched body of research, fundamental for every aspect of the discipline, evolutionarily intertwined with the demi-god Manhattan Project mathematicians who were in love with biology. Our patriarch is the guardian of the highest standards.

Santa Fe, the Manhattan Project, and John von Neumann intertwine in New Mexico and we can trace some evolutionary roots of computational biology from there. An evolutionary trajectory of theories, collaborations, and mentoring in this “land of enchantment” reaches to one common-ancestor node: John von Neumann, who proposed a research program aimed at a new computation and information theory essential for modeling the biological systems of the cell. The overarching goal of his new theory was the unification of continuous and discrete mathematics via the concept of statistical thermodynamic error. The two pillars to be unified, via statistical thermodynamics, were continuous mathematics, namely mathematical analysis, “the technically most successful and best elaborated part of mathematics,” and discrete mathematics, which “deals with rigid, all-or-none concepts, and has very little contact with the continuous concept of the real or complex number ... technically most refractory parts of mathematics ... by the nature of its approach, cut off from the best cultivated portions of mathematics, and forced onto the most difficult part of mathematical terrain, into combinatorics.” Professor Waterman’s work – with its pioneering influence, trailblazing new areas of research through a wide range of computational and mathematical methods applied to biological problems, weaving statistical and algorithmic methods towards a genuine statistics-powered continuous-discrete hybrid – is a singular body of work closest, as any other, to von Neumann’s herculean vision and research program.

I have known Professor Waterman for 20 years. Our collaboration began in 1992, when, two weeks after joining Sandia National Laboratories in Albuquerque, New Mexico, I participated in a Rutgers University tutorial workshop designed to attract computer scientists and mathematicians to work on the daunting research problems of the Human Genome Project begun by the U.S. Department of Energy (DOE). Professor Waterman was one of the workshop lecturers. At Sandia, I was in charge of establishing the Computational Biology Project, and I had been directed to work toward attracting Professor Waterman to Sandia by Dr. Fred Howes, the funding director of my new project, director of the Applied Mathematics Program at DOE (founded by von Neumann), and an enthusiastic supporter of computational biology. I succeeded, and Professor Waterman’s ensuing visits were instrumental in establishing our two-decade-long collaboration. Together with Professor Pavel Pevzner of University of California, San Diego, we three collaborated as leaders on a variety of projects providing computational and mathematical sciences foundations for computational biology and bioinformatics: the

RECOMB Conference, the Journal of Computational Biology, MIT Press Computational Molecular Biology book series, and Springer-Verlag Lecture Notes in Bioinformatics book series.

I remember vividly when Professor Waterman took Professor Pevzner and me in 1997 to the late Stan Ulam's house in Santa Fe, New Mexico, to bring flowers to his wife, Francois Ulam. It was an emotional visit, for it was as though we were making a connection to great mathematicians who preceded us. We were in Santa Fe to establish what has come to be known as the RECOMB conference – the Annual International Conference on Research in Computational Molecular Biology. It just so happened that von Neumann, involved in the war effort at Los Alamos, brought Ulam to New Mexico; Ulam brought Bill Beyer to New Mexico, Beyer brought Professor Waterman to New Mexico, and, thanks to Professor Waterman, we were in New Mexico for the first RECOMB conference. There, Professor Waterman passed the torch from mathematicians in love with biology at the end of the 20th century and the start of the 21st. Flanked by Bill Beyer and Nick Metropolis – two members of the greatest generation of mathematicians in residence in New Mexico – Professor Waterman declared: “This volume [the first RECOMB Proceedings] is proof positive of the vitality of a new discipline: computational biology.” The founding Steering Committee of the conference included, in addition to we three founders, Richard Karp (UC Berkeley), Tom Lengauer (Max Plank) and Ron Shamir (Tel-Aviv). Next year, the conference will celebrate its 17th year, and it truly has had an international journey: Santa Fe; New York City; Lyon, France; Tokyo; Montreal; Washington, D.C.; Berlin; San Diego, California; Cambridge, Massachusetts; Venice; San Francisco; Singapore; Tucson; Lisbon; Vancouver; Barcelona; Shanghai in 2013. The RECOMB conference is without question one of the top two international conferences in computational biology and bioinformatics, the other being the Intelligent Systems for Molecular Biology Conference. RECOMB owes a debt of gratitude to Professor Waterman, who back in 1997 was our flagship.

RECOMB, established as an Association for Computing Machinery conference, successfully borrowed for this exciting domain of computational biology the most effective traits of the exceedingly competitive top computer science conferences. RECOMB selects annually an internationally recognized program committee composed of the most active and distinguished researchers of the field, and applies the same program committee selection rules through refereeing and intense program committee team work. This effort, spanning many weeks, culminates in the selection of a very small fraction from the large number of high quality submissions. The RECOMB program is recognized for meeting a set of scientific standards that, for such a highly interdisciplinary field, require tremendous innovation and contributions. We have Professor Waterman to thank for this. He played a tremendous leadership role in all scientific aspects of the RECOMB conference, and has been a major contributor in designing the rules of scientific analysis that shape the advancement of the field by setting uncompromising standards for selecting invited speakers of highest distinction – with special attention to recognizing local pioneers in the host country – and ensuring fairness that is necessary in the confidential program committee work. Without question, Professor Waterman's contributions shaped the course of computational biology and bioinformatics towards the maturity of the field – a maturation that is reflected in RECOMB conference papers demonstrating significant innovation in computational and mathematical foundations of the field. The conference also provided a major educational component for graduate students, postdoctoral students, and young faculty about where the field is going by offering lectures by biologists of the highest reputation. The conference organizers are always making a major effort to secure funding for graduate students and postdoctoral students to attend in large numbers. The conference truly unified the field of computational biology and bioinformatics internationally, establishing RECOMB as the most prestigious level of achievement annually. Without question the 35 RECOMB papers accepted annually (selected usually from about 200-250 submissions) are the absolute top 35 papers of the field.

The Journal of Computational Biology is without question the top journal of the area devoted to innovation in statistical and computer science and mathematical methods. Professor Waterman's leadership of the Journal was instrumental in reaching this milestone. Undoubtedly, he is the best-selling author of textbooks on computational biology and bioinformatics. This reflects another shining dimension of his pedagogical philosophy and passion that resonated so well with generations of teachers of this new and exciting field: he is our Professor-in-Chief. His students hold positions of highest stature in computational biology, bioinformatics and biotechnology in many

countries – U.S., Germany, U.K., France and China, to name a few. In this respect, Professor Waterman has the most impressive record of all professors of the field.

Professor Waterman, together with Temple Smith, invented at Los Alamos National Laboratory the most beautiful algorithm of computational biology, the Smith-Waterman local alignment algorithm. I teach it in my Algorithmic Foundations of Computational Biology course at Brown, and it takes me about six lectures to cover the topic, although, as simplicity is the ultimate sophistication, it is a 3-lines algorithm. It is one of the deepest topics in computational biology; the intertwining of statistics and algorithms is truly von Neumannesque in harnessing the rigorous modeling of the fascinating but mysterious structure of the DNA sequence. The most cited paper in science is one describing the BLAST algorithm, with about 25,000 citations to date. Though Professor Waterman is not an author of that paper, the Smith-Waterman algorithm is at its heart. The ultimate tribute of a paper is that it is so much entrenched in the scientific culture that people do not cite it anymore. The Smith-Waterman part of BLAST is used every second of every day by thousands of users for 20+ years by now, and every genome assembly of every organism assembled to date, the human genome in particular, employs billions of runs of the algorithm. The Smith-Waterman and BLAST were essential for constructing the sequence of the human genome. All in all, the Smith-Waterman algorithm, with algorithmic speedups and exceedingly deep statistical theory, define “the practical” for computational biology methods and genomics analysis. Another topic of great depth that I teach in my class for about six lectures is the Lander-Waterman statistics – a most treasured topic for computational biology and genomics. Professor Waterman co-authored the gems of the technical discipline.

I co-authored only one paper with Professor Waterman, and it is very dear to me. I call it the “lighthouse” paper. I was then Senior Director of Informatics Research at Celera Genomics/Applied Biosystems, and Michael Hunkapiller, the president of the company, together with Craig Venter, my former boss and at that time president of what would become the Craig Venter Institute, asked me to lead the effort to build the tools for the most comprehensive comparison of all human genome assemblies to date, both of the public effort and of Celera. It was stressed to me that the paper describing this analysis needed to be most comprehensive and accurate, leaving no room for shifting conclusions and hopefully containing shining beams of success, like a lighthouse. Professor Waterman became a collaborator of ours at Celera and, together with a large group of colleagues, we published the lighthouse paper in the Proceedings of the National Academy of Sciences in 2004. Indeed, the results never moved since.

New Mexico is an enchanted land of algorithms and flowers, a safe haven for romantic liaisons between science and the arts where each can reach unprecedented peaks. It is here Professor Waterman found inspiration for writing literature. Los Alamos’ Nick Metropolis, whose algorithm is the most cited in the history of science and engineering, and Georgia O’Keeffe, whose paintings of desert flowers offer singular beauty, are characters in the professor’s piece, “Nick the Greek.” Within his prose you hear the rivers of his native Oregon, see the brightness of the light, and feel the freshness of the high-desert air of New Mexico that he calls skiing the sun.

Professor Waterman is without question the most influential educator and mentor of generations of computational biologists, and a model and an inspiration and mentor for many of us. He is the founder and leader of the top conference and top journal, and the author of the most respected textbook of the area. He is the co-inventor of the premier algorithm of genomics, a matter of national pride for the U.S. National Laboratories, having been developed at the same Los Alamos National Laboratory from which some of the most distinguished scientific work for the war effort emanated. In 2001, the computational construction of the sequence of the Human Genome, where his algorithm played a pivotal role, marked the phase transition point when biology became a quantitative science. The scientific and strategic importance of genomics for life sciences, medicine, agriculture, and government agencies led to the establishment of bioinformatics centers and institutes for quantitative biology all over the world. Due to the algorithm’s impact we witnessed the triumph of rigorous computational analytical methods in the Genome Era.

In sum, Professor Waterman is a national treasure! If I can add a personal thought, he brought us closer to von Neumann. Here are the lessons I learned from him, stated as “axioms” which are shared by von Neumann:

Axiom 1. Be a mathematician of both the continuous and discrete mathematics

Axiom 2. Be a world class scientist in at least one of the areas of your interdisciplinary research

Axiom 3. Be an intra-math, inter-sciences and cross-cultures scientist

Axiom 4. Be the guardian of high standards

Axiom 5. And in the end, the love you take is equal to the love you make

Sincerely,

A handwritten signature in black ink, appearing to read 'Sorin Istrail', with a stylized flourish at the end.

Sorin Istrail, Ph.D.

Julie Nguyen Brown Professor of Computational
and Mathematical Sciences

Professor of Computer Science

Editor-in-Chief, *Journal of Computational Biology*