CellPress

*Series: Training the next generation*

# Training the 21st Century Immunologist

**Roberto Spreafico[1], Simon Mitchell[1], and Alexander Hoffmann[1]**

Institute for Quantitative and Computational Biosciences, Boyer Hall, 611 Charles Young Drive, UCLA, Los Angeles, CA 90095

**Immunology, along with other fields of biology, is undergoing a revolution. Here we discuss the challenges and opportunities presented by considering the dynamical systems properties of the immune system, and harnessing the power of data-rich technologies. We present specific recommendations for changing graduate programs to incorporate training that will enable students to actively participate in the analyses of complex data and their biological system, and urge that we move from viewing quantitative and computational biology as interdisciplinary, to recognizing these as intrinsic to the discipline of immunology going forward.**

Biology is the study of highly complex dynamical systems. Indeed, at any scale - from eco-systems to populations, to organisms, to organs, to cells, molecular networks and macro-molecules - a hallmark of biological systems is the dynamical interplay of numerous components. It is remarkable how the tools of molecular biology and biochemistry have rendered this complexity tractable. Specifically, with the culmination of 'omic technologies, the molecular and cellular parts lists of cells are known, quantifiable, and increasingly readily available in electronic databases. This remarkable success at the same time signifies that biology has irreversibly changed to a data rich science.

This transformation, described more fully elsewhere (e.g., [1]), has changed what constitutes the skillset of a biologist. Up to recently, assay development often constituted a central aspect of training, as the ability to generate useful data was often limited by technical hurdles. With a growing number of cellular and molecular reagents, as well as highly sophisticated assays in kit form available from vendors, the ability to analyze data creatively and critically so as to obtain real insight becomes a distinguishing skill. Indeed, practices within our group reflect this change: whereas assay kits were banned in the initial years to ensure that students were trained in the skills of optimizing assays, our focus has shifted to requiring students to never take software output at face value, but to be able to customize data analysis methods. In other words, what distinguishes PhD biologists in the 21st century more so than previously, is innovation not in the generation of data, but in data analysis and interpretation.

Considering immunology specifically, two hallmarks stand out that guide our wish list for training of 21st century immunologists.

First, the immune system, as a distributed organ system involving networks of interacting cells throughout the body, is particularly rich in dynamical systems properties. Add the molecular networks in each cell, which show rich signaling and regulatory dynamics, and immunology may be considered the poster child of dynamical systems biology. Indeed, seminal studies in the field have uncovered fundamental immunological regulatory mechanism by involving mathematical dynamical-systems modeling. For example, through kinetic analysis and computational modeling, Ho *et al.* [2] identified a rapid turnover of HIV-1 *in vivo* that focused drug targets to protease inhibitors and greatly accelerated progress in HIV therapy. Pakker *et al.* [3] showed that the bi-phasic kinetics of memory T cell populations are mediated by redistribution of T cells in the blood rather than rapid proliferation. Savage *et al.* [4] used mathematical models to show that the T cell antigen receptor (TCR) repertoire selection can be attributed to the dissociation rate constants of TCR-peptide/MHC interactions, which are interpreted by a kinetic proof-reading mechanism. Our lab has followed an iterative approach of computational modeling and experimentation to understand the role of nested feedback loops to produce complex NFκB dynamics [5]. Thus, whether at the molecular or cellular network level, an understanding of how the immune system works in health and disease and how to devise therapeutic strategies, involves considerations of its dynamical systems properties. Indeed, mathematical models allow leveraging knowledge resulting from basic science and experimental model systems to interpret data-sparse clinical studies to enable not only the patient stratification but also predictive promise of 'precision medicine'.

What quantitative concepts and computational skills are then required to move forward our understanding of immunology? Biological function and phenotype are systems properties that emerge from the dynamical behavior of interacting networks. Students should be familiar with dynamical systems concepts, exemplified by immunological phenomena, such as dose-response behavior and ultrasensitivity, kinetic proof-reading, the diverse functions of negative and positive feedback loops in regulating both dose response curves and time series behavior, concepts of biomodality and bistability, thermodynamic stochasticity that determines intrinsic or extrinsic noise sources and network motifs that either tend to mitigate or amplify

*Corresponding author:* Hoffmann, A. (ahoffmann@ucla.edu).
[1]All authors contributed equally.

1471-4906/
 http://dx.doi.org/10.1016/j.it.2015.04.001

these. Further, PhD immunologists should be able to explore these concepts in the context of their own studies. To this end they need to acquire elemental skillsets in kinetic modeling and biophysics [6]. Ordinary differential equations provide the mathematical building blocks from which the dynamics of immunological systems can be represented and therefore a grasp of basic matrix algebra and calculus (limits of functions, differential equations, and the Euler method) is essential. Further, to enable the construction of kinetic models [7], the principles of computer programming (variables, functions, data structures, loops, and debugging) should be learned in a widely used programming language. MATLAB [8], Mathematica [9], and GNU Octave are widely used programming environments, while Berkeley Madonna [10], COPASI [11], and the MATLAB SimBiology toolbox offer more accessible, user-friendly interfaces.

The second hallmark of contemporary immunology is its harnessing of data-rich technologies, from next-generation sequencing (NGS), to mass cytometry, to imaging (both *in vitro* time-lapse and intravital two-photon microscopy), just to name a few. Time courses and single cell experiments, which are critical due to the dynamical and clonal nature of immune reactions, augment complexity by introducing additional dimensions. In fact, the realization that population measurements are often not informative has consolidated since immunologists pioneered flow cytometry to distinguish between the myriad of immune cell subsets. For similar reasons the ongoing transition of NGS analysis to single-cell resolution is particularly relevant to immunologists. After all, immune disorders often result from the misregulation of a few 'outliers', while the 'population mean' appears unaffected (e.g., [12]). Thus single cell analyses are critical for immunological research and compound the challenges of high-dimensional data analysis. Human studies add a further layer of complexity by introducing a much larger degree of between-subject variation than found in inbred animal models. Often variation represents only confounding noise that needs to be controlled for. But for certain questions, for instance in associations of genetic variants to phenotypic traits, variation is the very target of investigation. In either scenario, variation must be properly estimated at the experiment design stage, and must be properly modeled at the data mining stage.

To address the challenges of complex data analysis, immunologists require both conceptual understanding and working skillsets of statistical approaches. Concepts key to complex data analysis include the foundations of hypothesis testing such as difference and equivalence tests, a correct interpretation of *P*-values, the multiple testing problem and false discovery rates; elements of dimensionality reduction (such as principal component analysis and feature selection) and unsupervised clustering algorithms; regression and classification (including goodness of fit and ROC curves); as well as the foundations for experimental design such as statistical power analysis and factorial (full and fractional) design. In addition, immunology PhDs should have working or practical skillsets in manipulating the complex data they routinely generate. This involves proficiency with domain-specific software packages (such as FlowJo and CytoSPADE in flow cytometry), a plethora of NGS command-line tools organized in pipelines, but also statistically oriented computing environments such as R/Bioconductor that require basic facility with programming. These are critical to manipulate and integrate datasets, run statistical tests and to synthesize data into figures.

We do not advocate that every immunologist becomes an expert bioinformatician and computational systems biologist. Indeed, the depth and sophistication of these fields and tools, and their ongoing development, require that graduate education in bioinformatics and computational biology remain separate and be further expanded as well. However, every immunologist ought to have the knowledge and skills necessary to analyze and interpret the data they generate, and to relate it to relevant observations made by others, which are readily accessible in increasingly rich publicly available databases. Indeed, when immunologists are overwhelmed by the data that they have generated and hand over the task to computational scientists, valuable biological insight is lost. As such, the presently popular 'in-house experiment/outsourced analysis' workflow should be regarded as an emergency response that is far from optimal. It should be the domain of the immunologist to extract information from high-dimensional data by conjugating biological knowledge and analytical techniques. Indeed literacy in bioinformatics and computational systems biology is required for functioning as a 21$^{st}$ century immunologist – without it immunologists will find an increasing fraction of the literature unintelligible, unable to present in journal clubs, and down-the-line unable to participate in the review process. Further, an appreciation of the possibilities and challenges of quantitative and computational approaches, as well as shared vocabulary, is critical for engaging in productive collaborations with dry lab colleagues who could take the data analysis and systems modeling to another level.

How can we meet the demands of training the 21$^{st}$ century Immunologist? Here we offer five specific suggestions for immunology PhD programs.

First, we must recognize that changes in the skillsets demanded of immunologists also require different aptitudes. We suggest that admission criteria include proficiency in statistics and calculus as documented in classwork, as well as proficiency in a programming language. Indeed, we find that a common key hurdle to developing data analysis skills in graduate training is debilitating fear of programming. In an age where probably every middle school student is exposed to HTML and coding, this may function as a litmus test: some students took to it in school and developed programming skills further, and some did not. We argue that students who do not take to programming will always be hampered in their training as biological scientists.

Second, the first year core curriculum of graduate students should include statistics and dynamical systems courses that are taught by faculty in the program. Contemporary immunology includes statistical evaluation of large datasets and the analysis of the dynamical nature of the cellular and molecular networks. In other words, statistics and dynamical systems classes must focus on

immunological phenomena or risk being of little interest to immunology students [13]. Classes can take the form of survey lecture courses in which mathematical concepts are integrated, or a class of the contemporary literature, which invariably provide opportunities to examine the computational methods in detail. In an age where knowledge and facts, even sophisticated immunology, is at everyone's fingertips, fact-focused classes may give way to such quantitative-concept-focused classes that benefit from problem sets and homework.

Third, practical quantitative and computational analysis skills should be trained in computational laboratory courses. These courses should include lectures in which tools and techniques are introduced, and homework tasks in which graduate students attempt to apply what has been taught to gain insight into a real example in immunology. The classes should be taught by practitioners of computational biology: teaching assistants must be systems biologists and bioinformaticians, and lab courses may be team-taught by postdocs from immunology labs. The methods used in the homework and insight gained should be briefly presented to the group by the graduate students at the next class: this approach has proven success in introducing quantitative skills to bioscientists [14]. There is a trend to attempt to gain practical skills rapidly at the start of graduate school through intensive 'bootcamps' [15], but these may not allow the required time for homework, hands-on learning, independent attempts at overcoming pitfalls and extracting biological insight from computational studies.

Fourth, the qualifying exam must reflect the emphasis in quantitative and computational understanding and skills in graduate training. A typical format for such an exam is a research proposal unrelated to the student's ongoing research, and such a format is amenable to this goal. By requiring preliminary results based on the analysis of publicly available datasets and/or kinetic models, students will be required to demonstrate a working knowledge and the ability to defend their analyses, choice of statistical tests, or parameter values in the oral portion of the exam. Having demonstrated these proficiencies, students will be on track to become 21st century immunologists who harness the opportunities of large and complex datasets and push our understanding of the complex dynamical immune system forward.

Fifth, immunology programs should provide opportunities for continued professional development. Software tools available change rapidly, requiring periodic retraining of immunologists not only at the graduate student, but also at the postdoctoral fellow, research assistant, and faculty level. Massive open online courses (MOOCs) provide such opportunities. In addition, the traditional classroom settings can be effective. At UCLA's QCB Collaboratory, we have found that a useful format is the week-long workshop of two or three hours daily focused on specific hands-on skillsets at the computer. Homework tasks again are an important component in conveying the skills. These workshops may be effectively taught by practitioners in the field (postdoctoral fellows) and respond to demand or the availability of important new software packages.

Thus, transforming the immunology graduate program to meet the challenges and opportunities of 21st century immunology appears to have a surprisingly straightforward recipe. The key challenge seems to lie not with the availability of appropriate students, but of faculty, who can assess applications, teach the courses outlined above, and evaluate qualifying exams. Ongoing professional development may help but recruitment of new faculty or suitable faculty from related disciplines into immunology programs is critical. This represents a broadly accepted wisdom in the transformation of the biosciences, but we urge that we go further: rather than viewing quantitative and computational biology as being interdisciplinary, we view them as intrinsic parts of the biological discipline going forward. Computational biologists are biologists as much as molecular or cell biologists are. After all, biology in the 21st century is arguably the most data-rich science of the most intricately regulated dynamical systems that any discipline has to offer. Further, computational approaches offer means to leverage the knowledge gained from the basic science enterprise to deliver on the promise of personalized, precision medicine. The faculty of biology departments ought to reflect this fact.

## Acknowledgments

R.S. and S.M. thank Matteo Pellegrini and the QCB Collaboratory for the valuable experience of teaching quantitative skillsets to a broad range of researchers and students.

## References

1 National Research Council Committee on Undergraduate Biology Education to Prepare Research Scientists for the 21st Century, (2003) BIO2010: Transforming undergraduate education for future research biologists. National Academies Press
2 Ho, D.D. *et al.* (1995) Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. *Nature* 373, 123–126
3 Pakker, N.G. *et al.* (1998) Biphasic kinetics of peripheral blood T cells after triple combination therapy in HIV-1 infection: a composite of redistribution and proliferation. *Nat. Med.* 4, 208–214
4 Savage, P.A. *et al.* (1999) A kinetic basis for T cell receptor repertoire selection during an immune response. *Immunity* 10, 485–492
5 Basak, S. *et al.* (2012) Lessons from mathematically modeling the NF-kappaB pathway. *Immunol. Rev.* 246, 221–238
6 Vallabhajosyula, R.R. and Raval, A. (2010) Computational modeling in systems biology. *Methods Mol. Biol.* 662, 97–120
7 Mitchell, S. *et al.* (2015) Studying NF-kappaB Signaling with Mathematical Models. *Methods Mol. Biol.* 1280, 647–661
8 *MATLAB and Statistics Toolbox Release* 2012b, The MathWorks Inc.: Natick, MA, United States
9 *Mathematica Version 6.0*. 2007, Wolfram Research Inc.: Champaign, IL, United States
10 Macey, R. *et al.* (2000) *Berkeley Madonna user's guide*, University of California, (Berkeley, CA)
11 Hoops, S. *et al.* (2006) COPASI—a complex pathway simulator. *Bioinformatics* 22, 3067–3074
12 Spreafico, R. *et al.* (2014) A circulating reservoir of pathogenic-like CD4+ T cells shares a genetic and phenotypic signature with the inflamed synovial micro-environment. *Ann. Rheum. Dis.* Published online December 12, 2014. http://dx.doi.org/10.1136/annrheumdis-2014-206226
13 Redish, E.F. and Cooke, T.J. (2013) *Learning each other's ropes: negotiating interdisciplinary authenticity*. *CBE Life Sci. Educ.* 12, 175–186
14 Potter, W. *et al.* (2014) Sixteen years of collaborative learning through active sense-making in physics (CLASP) at UC Davis. *Am. J. Phys.* 82, 153–163
15 Schatz, M.C. (2012) Computational thinking in the era of big data biology. *Genome Biol.* 13, 177