Written evidence submitted by British Society for Immunology

Introduction

1.1. The British Society for Immunology (BSI) is the largest immunology society in Europe. We represent the interests of over 3000 immunologists working in academia, clinical medicine, and industry. Our main objective is to promote and support excellence in research, scholarship and clinical practice in immunology for the benefit of human and animal health.

1.2. As a nation we are world leading in our immunological research and rank first for research in infection and immunology in the G7.[i] Immunological science underpins many aspects of human health and the progression of disease. The application of immunological research extends across communicable disease and vaccination to the management and treatment of chronic diseases such as diabetes, asthma, allergies, and even cancer. It is also now becoming clear that immune responses are key to the development of many common disorders not traditionally viewed as immunologic, including metabolic, cardiovascular, and neurodegenerative conditions. These diseases are growing in prevalence and represent significant health challenges for the 21st Century.

Summary

1.1. We welcome the opportunity to submit evidence to this inquiry, where we will focus on the acute skills shortage in computational biology and bioinformatics. Driven by the expansion of affordable next-generation sequencing technologies, deficiencies in computational and bioinformatics skills have led to an interpretation and analysis bottleneck which impedes the adoption of high-throughput technology in informing research on complex issues in immunology, such as in vaccines research and the development of new cancer immunotherapies.

1.2. Particularly in demand are skills such informatics analysis, algorithm research, software development, and scripting in common programming languages including R, Perl, and Python. Attempts to address the undersupply of these skills have largely involved the input of individuals from numerical backgrounds (maths, computing science, physics). However, we believe there is a strong need to develop this experience within the life sciences workforce itself, including in immunology, to blend the quantitative understanding required to interpret large molecular datasets with the intuition of deep biomedical experience.

1.3. Skilling up in this area will require action at both undergraduate and postgraduate levels, including through the integration of non-elective courses and seminars on systems biology, informatics, and big data science into taught degree programmes. For the development of more advanced capability at post-doctoral level we would like to highlight the Computational Genomics Analysis and Training programme within the MRC Weatherall Institute of Molecular Medicine at the University of Oxford as an example of the type of model which could be replicated to expand provision of analysis and training in this area. Representatives from the programme would be happy to provider further information or to be contacted by the Committee if required.

Computational and quantitative skills in the life sciences

3.1. The immune system is an extraordinarily complex system of organs, tissues and cells. Understanding it is an immense challenge. Since the days of Jenner in the late 18th Century, classical, experimental immunology has driven advances in our understanding of immune function and disease pathogenesis. In recent decades however the emergence of affordable next-generation sequencing technology has moved significant elements of immunological research into a new sphere, where computational and mathematical modelling of molecular systems enables researchers to investigate the complexity of the immune system in unprecedented fidelity.

3.2. Computational systems biology, driven by the "omics" technologies (single cell transcriptomics, genomics, proteomics etc), blends experimental and computational research to better understand complex systems in a holistic way (e.g. proteomics assesses changes in protein expression across the entirety of a specific biological context, for example a cell). Computational biology is often linked with bioinformatics, which uses statistical methods (often involving the utilisation of tools and software created through computer programming) to model and simulate complex systems. These techniques are an emerging though increasingly prominent aspect of many research activities. Their utility in immunology is broad but includes the following examples:

3.2.1. Vaccine development is traditionally based on an empirical approach. This method is both time consuming and costly, as many vaccine candidates ultimately fail to make it through clinical trials. Deploying a systems biology approach to better understand the molecular processes behind a candidate molecule's immunogenicity and safety profile, and using this data to inform early development, helps accelerate and improve the development of new vaccines.[ii]

3.2.2. Immune suppression is a central hallmark of cancer. The sequencing of cancerous tumours helps build a better picture of the genetic basis of cancer and how the disease interacts with the immune system, opening up new avenues for therapeutic intervention. Mining this big data to identify the expression of cancer specific antigens is important in the development of new cancer immunotherapies.

3.3. Together, these approaches produce and exploit huge datasets, driving the "big data" revolution in immunology. The emergence and continued expansion of big data science, systems thinking, and immunoinformatics has created an acute and widely acknowledged demand for researchers with the computational and quantitative skills required to both generate and interrogate this information in a sophisticated way.[iii, iv]The undersupply of these skills in immunology (and the broader life sciences) means that researchers often rely on the expertise of scientists from different backgrounds, notably physics and applied mathematics. While the contribution of these individuals and the skills and experience they have brought into biology has been much needed, we would argue that there is strong demand to develop these skills within the immunology workforce itself.

3.4. Skilling up in this area, and therefore closing the divide that exists between the service (i.e. those with computational and mathematical skills) and immunology (those with wet lab skills and the immunological expertise to explore research questions from the data) has clear

advantages. For example, being able to generate the data is one thing, having the confidence to then use it to draw firm biological observations and validate these through experimental observation quite another. While not every immunologist is required to become a computer scientist per se, developing experience within the workforce in informatics analysis, algorithm research, software development, and scripting in languages such as R, Perl, and Python, would help address the bottleneck in analysis and interpretation that currently exists in immunology and other disciplines.

3.5.   Importantly the demand for these skills is neither unique to immunology nor the UK. Domestically, sectors such as banking, insurance and tech compete for the same skillset, exerting further pressure on the skills gap. These skills are also in high demand on the global labour market and many countries, notably China (particularly through large international institutes like the Beijing Genomics Institute)[v], specifically recruit international candidates to fill vacant roles in computational biology. While the departure of talented scientists from the UK to countries like China can be seen as a loss, the export of skills and expertise may also be an opportunity to enhance and strengthen scientific collaboration (see recommendation 3 of the BSI's report Immunology: An international, life-saving science[vi]).

3.6.   Moreover, addressing these skill gaps must also be considered within the context of the Government's Industrial Strategy. Innovation intensive industries are a key strength of the UK. Nevertheless, there is significant scope to add further value to these industries (including life sciences and pharmaceuticals) by addressing key skills shortages. Indeed, three quarters of roles on the Home Office Shortage Occupation List are STEM roles (bio-informaticians are included).[vii] Meeting the skills gap must be seen as fundamental to the success of the UK as a thriving knowledge-based economy.

Addressing this skills shortage

Undergraduate

4.1.   Addressing the shortage in computational and bioinformatics skills requires action at both the undergraduate and postgraduate levels. Undergraduates would benefit from integration of teaching on big data science, systems biology, genomics, and bioinformatics into life science degree programmes. This should ideally form a core component of the students' degree (i.e. not elective) and could include cross-faculty initiatives with colleagues in computer science and maths.[viii] Significantly, with the revision of the national computing curriculum and the inclusion of coding into the curriculum, students entering higher education are increasingly computer literate. The goal of such undergraduate training should therefore be to instil in young scientists a familiarity and understanding of the techniques and approaches that form the interface of biomedical research and big data science. This understanding can then form the basis of more advanced learning at postgraduate level.

Postgraduate and beyond: MRC Computational Genomics Analysis and Training model

4.2.   Many universities now offer taught postgraduate degrees, such as an MSc, in bioinformatics or systems biology. Whilst these programmes are a useful opportunity for interested graduates to further their training in bioinformatics, we believe there is significant scope for increasing training opportunities for postgraduates and postdoctoral students, particularly in the context of longer form training which mixes tuition with real world research and application. One such

model in operation is the MRC funded Computational Genomics Analysis and Training (CGAT) programme at the MRC Weatherall Institute of Molecular Medicine, based at the University of Oxford.[ix] CGAT's mission is to train future leaders in computational genomics in order to address the acute shortage of skills and capacity in the interpretation of high-throughput genomics. Its model involves taking on Genomic Training Fellows at the postdoctoral level from a variety of backgrounds, such as immunology, genetics, and molecular biology, for in-depth training in all aspects of computational genomics for a period of three years.

4.3. The CGAT programme includes an initial assessment period to assess existing computational skills before fellows embark on a tailored training programme. Once this initial period is complete, fellows then have the opportunity to apply their learning in a real world context through collaboration with UK-based research groups. This model not only ensures provision of much needed analysis capability to research scientists throughout the UK, but also has the benefit of enabling fellows to take the lead on computational analysis and interpretation of complex biological questions. In recognition of their input, fellows are awarded joint first authorship on publications as recognition of the value of their contribution. At the end of their fellowship, CGAT fellows return to their discipline as leaders in computational analysis, becoming important ambassadorial alumni with the expertise to train others in the same skills (for more information on the CGAT model see Sims et al *CGAT: a model for immersive personalised training in computational genomics*[x]).

4.4. We believe that this model and its principles – taking numerate biologists, training them in more advanced bioinformatics, enabling them to apply this expertise in the real world, and developing in them the independence to become computational leaders in their own field – is highly franchisable, and with appropriate recognition and support could well be replicated elsewhere to expand the provision of much needed computational analysis in immunology and other disciplines. Interest in and competition for CGAT fellowships is intense, with around 100 applications for each place in a recruitment round, underlining the scale of demand for more training capacity in this area. For more information, representatives from the programme would be more than happy to input further into the inquiry.

References

[i] APPG on Global Health (2015). The UK's Contribution to Health Globally.
[ii] Blohmke, C. (2015). The use of systems biology and immunological big data to guide vaccine development.
[iii] ABPI (2015). Bridging the skills gap in the biopharmaceutical industry.
[iv] Tetradata research project. Data science in Europe – skills gap or emerging talent?
[v] Beijing Genomics Institute
[vi] British Society for Immunology. 2016). Immunology: An international, life-saving science.
[vii] Home Affairs Committee (2015). Fifth Report of Session 2015-16. Immigration: skill shortages.
[viii] Schultze, J (2015). Teaching 'big data' analysis to young immunologists.
[ix] Computational Genomics Analysis and Training (CGAT)
[x] Sims et al (2015). CGAT: a model for immersive personalised training in computational genomics.