

Graduate Training at the Interface of Computational and Experimental Biology: An Outcome Report from a Partnership of Volunteers between a University and a National Laboratory

Albrecht G. von Arnim^{1,2*} and Anamika Missra¹

¹Graduate School of Genome Science and Technology, University of Tennessee Oak Ridge National Laboratory, and ²Department of Biochemistry and Cellular & Molecular Biology, University of Tennessee, Knoxville, TN 37996

ABSTRACT

Leading voices in the biological sciences have called for a transformation in graduate education leading to the PhD degree. One area commonly singled out for growth and innovation is cross-training in computational science. In 1998, the University of Tennessee (UT) founded an intercollegiate graduate program called the UT-ORNL Graduate School of Genome Science and Technology in partnership with the nearby Oak Ridge National Laboratory. Here, we report outcome data that attest to the program's effectiveness in graduating computationally enabled biologists for diverse careers. Among 77 PhD graduates since 2003, the majority came with traditional degrees in the biological sciences, yet two-thirds moved into computational or hybrid (computational–experimental) positions. We describe the curriculum of the program and how it has changed. We also summarize how the program seeks to establish cohesion between computational and experimental biologists. This type of program can respond flexibly and dynamically to unmet training needs. In conclusion, this study from a flagship, state-supported university may serve as a reference point for creating a stable, degree-granting, interdepartmental graduate program in computational biology and allied areas.

INTRODUCTION

Over the past 20 years, the traditional subdisciplines of the biological sciences have become more integrated, and this integration has allowed biologists to tackle grand challenges in biomedical, agricultural, and environmental research. The collaborative work espoused by 21st-century biology relies increasingly on large and multidimensional data sets that require the close collaboration of experimental biologists and computational scientists. The U.S. Bureau of Labor Statistics projects additional job growth in the life and physical sciences, and biomedical engineers (Lacey and Wright, 2009) and statisticians (Hogan and Roberts, 2015) are two of the faster-growing occupations. Because both the biological systems under study and the novel technologies used to analyze them are exceedingly complex, a new type of biologist is needed, who has an active knowledge of both biological as well as computational techniques, and who can communicate effectively with scientists whose expertise is more restricted to either wet-lab or dry-lab environments. In his 2012 editorial for a special issue on computational biology, the editor in chief of *Science* magazine, Bruce Alberts, wrote, “Graduate schools have become increasingly enthusiastic about recruiting outstanding students with strong backgrounds in (the physical and computational sciences) to address challenging problems in the biological and biomedical sciences. This is an important trend, except that too often faculty mistakenly assume that learning biology

Erin L. Dolan, *Monitoring Editor*

Submitted February 21, 2017; Revised July 26, 2017; Accepted July 27, 2017

CBE Life Sci Educ December 1, 2017 16:ar61

DOI:10.1187/cbe.17-02-0038

Conflict of interest statement: The corresponding author has served as the director of the program being described here since 2010.

*Address correspondence to: Albrecht G. von Arnim (vonarnim@utk.edu).

© 2017 A. G. von Arnim and A. Missra. CBE—Life Sciences Education © 2017 The American Society for Cell Biology. This article is distributed by The American Society for Cell Biology under license from the author(s). It is available to the public under an Attribution–Noncommercial–Share Alike 3.0 Unported Creative Commons License (<http://creativecommons.org/licenses/by-nc-sa/3.0>).

“ASCB®” and “The American Society for Cell Biology®” are registered trademarks of The American Society for Cell Biology.

is easy leaving these very talented young people nearly on their own to acquire the biological wisdom that they will need to explore the many mysteries of living systems” (Alberts, 2012).

Accordingly, many universities did indeed create graduate programs in quantitative biology, computational biology, or bioengineering (e.g., Noble *et al.*, 2016). Meanwhile, marking a second and independent trend, it has been called to our attention that many PhD-level scientists do not continue in research-intensive careers, most do not enter into academic faculty positions, and not all faculty positions are research intensive (Griffiths *et al.*, 1995; Kennedy *et al.*, 2004; National Research Council, 2011; Tilghman and Rockey, 2012; Bourne, 2013; National Science Board, 2014; National Science Foundation [NSF], 2014b,c; Daniels, 2015; Leshner, 2015; Gibbs, 2016). The career aspirations of graduate students also change; several studies indicate a trend away from research-intensive academic faculty careers (Sauermann and Roach, 2012; Gibbs and Griffin, 2013; Woolston, 2015; Layton *et al.*, 2016). These data have given rise to the concept of a branching network of career development pathways (Fuhrmann *et al.*, 2011; Gibbs and Marsteller, 2016; Layton *et al.*, 2016; St. Clair *et al.*, 2017). Senior researchers (Yamamoto, 2014; Leshner, 2015) and funding agencies (Singh *et al.*, 2015; Gibbs, 2016; Gammie *et al.*, 2017) in the life sciences have called for a transformation of how PhD students are trained, and not only recently (Griffiths *et al.*, 1995), advocating for additional emphasis on transferable skills, that is, skills that will hold value even for students who leave original research, to complement the research skills that lie at the center of most PhD programs. Similar or complementary perspectives have also been articulated by interest groups representing PhD students and postdocs (Johnson, 2014; McDowell *et al.*, 2014; Teitelbaum, 2014; Polka *et al.*, 2015).

The School of Genome Science and Technology (the GST program) at the University of Tennessee (UT) has trained graduate students at the nexus of experimental and computational biology for 18 years. GST is a cross-disciplinary, degree-granting, graduate program in the life sciences that links the campuses of Oak Ridge National Laboratory (ORNL) and the UT. The partnership between a university and a national laboratory provides a mechanism for scientists from ORNL to contribute to graduate training and an opportunity for graduate students to participate in the scientific culture of a government lab. The program strives to contribute to a transformation in how PhD students are being trained. It has implemented some but not all of the aspirations expressed by leaders and vocal clients of graduate training. The program enthusiastically endorses the concept that graduate students ought to be trained for diverse careers, including research-related scientific careers and careers in industry. The program is also distinct from almost any program that awarded the PhD degrees of the current GST faculty.

Here, we describe how the program was originally conceived, and how it has changed, serving evidence that a stand-alone, intercollegiate program can adapt to new challenges in a flexible manner. We describe the administrative structure, institutional support, and curriculum of the program, including our efforts to 1) introduce all students to computational biology in a life science context and 2) confer transferable skills such as communication, team science, peer mentoring, and peer learning and instill among students a sense of cohesion and the moti-

vation to become stewards of their discipline. Despite calls to make the outcomes of graduate programs more easily accessible (Singh *et al.*, 2015; Gibbs, 2016), few PhD programs do so. Arguably, few if any outcome studies on life science PhD programs have been published (Kuehnle *et al.*, 2009; University of California–San Francisco and Graduate Division, 2013; Brokaw and O’Loughlin, 2015; Polka *et al.*, 2015; Noble *et al.*, 2016). We present and discuss our outcome data, many of which are publicly accessible, to allow an assessment of the degree to which the program has succeeded in its goal of training computationally enabled biologists and hybrid experimental–computational biologists for diverse careers in academic, industrial, and governmental laboratories. Our data may serve as a reference point for similar programs, current or future, hosted on state-supported university campuses elsewhere.

METHODS

This study reports data that were collected by the office of the GST director since the inception of the program in 1998. The data are presented primarily in the form of summary statistics. Comparisons between groups were based on Fisher’s exact test, two-sided. The authors of the study serve as the management team of the program. The study was examined by the UT Institutional Review Board and declared to be exempt from review (UTK IRB-17-03540-XM).

Participants

Data were collected and reported on two subsets of GST students—those recruited to the program between 2010 and 2016 and PhD alumni from 2003 to 2016. The data on recruited students include students who completed an MS degree or a PhD degree, left the program, or remain in the program. For comparison purposes, data on recruits and alumni were also collected and reported for another life science graduate program at the UT—Biochemistry and Cellular & Molecular Biology (BCMB). The demographic data were similar for both programs (Table 1).

Data Collection and Analysis

Data on career trajectories of alumni were obtained from the LinkedIn social media network, from institutional websites, from alumni’s publications, and from personal communications

TABLE 1. Demographic information for GST and BCMB alumni from 2003 to 2016 and recruits from 2010 to 2016

	GST		BCMB	
	Alumni	Recruits	Alumni	Recruits
Gender				
Male	49	32	41	34
Female	28	25	24	36
International	42	30 ^a	34	22
Race/ethnicity				
Asian	31	24	34	19
African American	5	3	1	2
Hispanic	0	3	0	6
Caucasian	41	27	30	43
Total	77	57	65	70

^aThis number is significantly elevated as compared with the other program (Fisher’s exact test with $p < 0.05$).

between alumni and GST staff. Multiple sources were needed, because each source was incomplete. The data for each of the alumni were collected into a spreadsheet in February 2017 with the goal of populating the cells for first, second, and any subsequent positions until the current position, as applicable. To minimize errors from the heterogeneity of data sources, we cross-checked the information against other sources whenever possible, and few discrepancies were noted. The classifications were performed by the authors. When classifying the diverse professional paths of our alumni, we distinguished between the first position after the PhD degree (usually but not always postdoctoral research), and one later (“secondary”) position. If the first position was a brief period of employment in the PhD adviser’s group of less than 6 months, this position was ignored. It is not possible to ascertain whether a given position is “stable,” “long-term,” or “permanent” based solely on the job title or the perception of the person occupying it, and we therefore did not qualify positions in this way. We classified positions as research (postdoc, research staff, support staff, research faculty, product development in industry, etc.), research-related science (postsecondary teaching, intellectual property, publishing), science-related (consulting, K–12 teaching, pharmacy), and non-science (e.g., finance, healthcare, Web services). We do not report unemployment, but based on available records, the fraction of GST alumni who are voluntarily outside the workforce or otherwise unemployed is below 5%. For the first position, positions titled “postdoctoral associate” or “research associate” were classified as “postdoctoral research,” while positions titled “staff scientist,” “bioinformaticist,” “computational scientist,” “instructor,” and others were not considered to be postdocs. Positions in postsecondary instruction at primarily undergraduate colleges were considered “academic teaching.” Positions at research hospitals were classified as “hospital,” while positions in medical schools were “academic.” Positions at national labs such as ORNL, National Institutes of Health (NIH), Food and Drug Administration, or the Environmental Protection Agency were classified as “governmental.”

Data on matriculated students were obtained from admission files and were supplemented or cross-checked with personal communications. For prior degrees, when students reported multiple degrees or majors, only the more advanced degree was considered, and only the major most related to life sciences. To classify the type of dissertation research, wet-lab/experimental, dry-lab/computational, or hybrid, we relied on dissertations and on the authors’ personal recollections of the alumni’s research area, as well as LinkedIn data. The research was considered “hybrid” only if the experimental component was substantial and if the computational component involved quantitative analytical skills that go beyond those common among life scientists.

Learning Goals of the Program

The GST program seeks to train life scientists across the intersection of wet-lab (experimental) and dry-lab (computational) biology. The term “wet-lab” refers to hands-on research with actual molecules, cells, or organisms, while “dry-lab” refers to research with data resident on a computer, without directly handling biological specimens. For the purpose of reporting to the university’s accreditation agency, the Southern Association of Colleges and Schools, GST reports five sets of learning goals. These include 1) analytical thinking, 2) research compe-

tence, 3) teamwork and professional development, 4) teaching, and 5) communication skills, for which numerical data can be gathered that serve as metrics for the value that the student body overall derives from this program (metrics are listed in Supplemental Material). However, this article aims to report on the professional career outcomes of GST’s alumni, not whether the GST program is meeting its learning goals.

Specific learning goals such as teamwork that are emphasized in GST build on specific features of the program; for example, scientific projects at ORNL often involve larger teams than at a university. Likewise, the general directive that students develop research competence in computational and experimental research is supported by the course curriculum as well as by recruiting and advising by the program. For example, students are guided toward courses to learn command line programming and the quantitative foundations of various analytical biochemistry and “omics” techniques. The Supplemental Material includes selected course syllabi as well as a summary of the perceived benefits and challenges of the GST model and additional perceptions of the institutional changes brought about by the program.

IMPLEMENTATION

Administrative Structure of the Graduate Program

The GST program is a degree-granting graduate program at the flagship campus of the UT that is administered in the College of Arts and Sciences. The program trains ~40–50 graduate students at any given time, at least 80% of whom are in the PhD track. GST is not a department and does not have faculty positions. The primary advisers of GST students are scientists with faculty appointments in any one of five colleges at UT or staff scientists at the ORNL, which is located 30 miles from the UT campus. Scientists who qualify and apply for membership in GST are granted adjunct status in GST by the provost after a vote by the GST faculty. The director of the GST program holds a regular faculty appointment in Arts and Sciences with a small administrative supplement in GST. The director is assisted by two administrative staff members and a series of committees composed of GST faculty members from UT and ORNL.

The financial support for GST students comes from graduate teaching assistantships, external research grants or ORNL projects, external fellowships, and a predoctoral training grant dedicated to female and underrepresented minority students funded by the NIH. First-year students in GST have been supported by funds raised from ORNL and UT, rather than faculty research grants, because first-year students follow an intensive curriculum of lab rotations and core courses with a broad and cross-disciplinary scope (see *Curriculum*).

Historical Development of the Program’s Computational Biology Portfolio

The GST program was conceived and launched in its current form in 1998 to develop a synergy in research training between the life sciences at the UT–Knoxville, the land-grant campus of the State of Tennessee, and ORNL. At the time, with genome sequencing in full swing, the venerable mouse genetics program (“Mouse House”) at ORNL was being catapulted into the genome age. Computational biology was perceived as a cornerstone in interpreting genomics data, and with foresight, mammalian and yeast geneticists closed ranks with a small group of computational scientists to launch the GST program.

GST's portfolio has been remarkably dynamic over the 18 years of the program. As ORNL dismantled the mouse genetics group and became the lead institution of the Department of Energy-funded Bioenergy Science Center, the GST program began to serve the research communities of microbial ecology and plant genetics and genomics. On the UT campus, GST recruited additional faculty members from the Colleges of Agriculture, Veterinary Medicine, and elsewhere. The guiding principle of faculty recruitment to GST is to build a community of researchers with an interest in quantitative and computational biology. The growth of GST into a substantive computational training environment was augmented by the installation of several supercomputers at ORNL, which included what were for a period of time the world's fastest academic computers, named Kraken and Titan. The computational expertise within the program includes both bioinformatics and mathematical modeling; the subject areas range from the molecular dynamics simulation of proteins at the atomic scale all the way to evolutionary biology at the genome scale. At this moment, approximately 16 GST faculty members are computational biologists, that is, developers of algorithms, databases, mathematical models, bioinformatic analysis pipelines, and high-performance computing workflows. Yet the majority of GST's more than 70 faculty members are experimentalists and as such are users as well as conveyors of computational expertise.

Curriculum

Although the program was not specifically conceived as a computationally focused program, it is now the central mission of GST to train researchers with both biological and computational expertise to conduct investigations at the nexus of computational and experimental biology. We emphasize this direction in our recruiting, and our applicants have taken notice—a plurality of them express an interest in this area. The program is rooted in the life sciences, focused on the hypothesis-driven investigation of natural processes rather than solving computational or engineering problems. Briefly, GST comprises a highly structured core curriculum in the first year, a more individualized second year that culminates in the comprehensive exam,

and a third phase that focuses increasingly on dissertation research while also engaging students to develop other transferable professional skills. As is typical for PhD programs in the life sciences, students allot most of their time to dissertation research, followed in approximate order by course work, teaching, and other professional development, such as diverse forms of service, peer mentoring, and outreach. Thus, the activities of students over the course of the program closely align with its overarching learning goals (see the Supplemental Material).

The First Year. Students pursue up to four research rotations with the goal of honing their experimental skills and identifying a dissertation mentor. The rotations are generally too short to produce tangible scientific conclusions, but they sometimes do. Meanwhile, a series of classroom courses provides background in molecular genetics, protein biochemistry, and computer programming. GST students take more courses than students in our sister programs, because, first, in the spirit of interdisciplinary training, a solid foundation and broad introduction to diverse areas of the life sciences is essential. We envisage that GST graduates will be the ones building the bridges in their future teams. Second, most students come to GST with a certain degree of specialization (e.g., molecular genetics, bioinformatics, environmental biotechnology) or with a non-life science degree (e.g., physics, chemical engineering, computer science) and want to develop a foundation in life sciences.

The core courses in the first year include the following (Table 2):

1. A series of two semester-long courses in genetics and genomics (GST I: microbial, plant, and biomedical science) and analytical technologies (GST II: mass spectrometry and structural biology; see the Supplemental Material for course syllabi). The segment on plant genetics in particular emphasizes transferable skills, such as writing a short research proposal, and occasional breakout sessions for peer-to-peer discussion in class. We also discuss opinion papers and commentaries on project design and career development (Alon, 2009) and

TABLE 2. GST core courses

Number and title	Format	Topics
LFSC520 GST-I Genetics & Genomics	Lecture and projects	Microbial and plant genetics and genomics; cancer cell biology; population genetics
LFSC521 GST-II Analytical Technologies	Lecture and projects	Mass spectrometry; nuclear magnetic resonance; x-ray crystallography; biophysical chemistry; next-generation sequencing
LFSC507 Computer Programming	Computer lab	Programming; Linux; Python; graphing and biostatistics
LFSC517 Comparative Genomics	Lecture	Molecular evolution; comparative genomics
BCMB511 Advanced Protein Biochemistry	Lecture and projects	Enzymology; membrane protein structure and function; protein trafficking; cytoskeleton
BCMB512 Advanced Molecular Biology	Lecture and projects	Gene regulation (chromatin, RNA processing, translation); cell cycle and cytokinesis
LFSC541 Colloquium	Presentations, projects	Research presentations; tutorials; invited speakers
LFSC515 Introduction to GST	Presentations, discussion	Orientation upon entry into the program; GST faculty members introduce research

interviews with leading scientists (Petsko, 2003; Benfey, 2013).

- Two courses in computational biology. The first of these teaches computer programming to naive students who lack any background in this area (LFSC507: Linux command line, Python, and biostatistical applications). The second course (LFSC517) covers molecular evolution and comparative genomics. Alternatively, or in addition, students may choose a course that teaches programming in C++, which is offered in the Computer Science Department (COS505).
- A series of two semester-long courses in advanced protein biochemistry (BCMB511) and advanced molecular biology (BCMB512). These courses are challenging, and because GST students take them as a group, many of them form a cohesive group and also engage in some friendly competition with students from other programs (Biochemistry, Microbiology). One of these courses may be waived if students have a strong background in molecular biology or are clearly not leaning toward protein biochemistry.

The course work primarily supports learning goals 1, analytical thinking; and 2, research competence; as well as 3, teamwork and professional development; and 5, communication.

The Second Year. In the second year, students ramp up their dissertation research. Besides this, they typically take a course in grant writing, strongly recommended as preparation for the comprehensive exam. Students are also encouraged to advance their computational skills. There are many options for this. For example, UT offers an Interdisciplinary Graduate Minor in Computational Science, which consists of 15 credits of coursework, six to nine of which are already completed during the core curriculum. The minor asks students to take an additional six to nine credits in mathematics or statistics and in computer science. We recently examined whether the minor adds time to degree. The 10 PhD students in GST who completed the minor graduated within an average of 5.3 years, no later than the average of 5.5 years for GST students overall. Most students also serve as teaching assistants during their second year, a condition of their financial support from the GST program. In addition, the GST curriculum committee instituted that all PhD students should gain experience in teaching (see learning goal 4, teaching); and indeed, many GST students later assume positions with substantial teaching responsibilities (see *GST Alumni Pursue Diverse Career Paths*).

The Third Year. In the Fall semester of the third year, GST students tackle the comprehensive examination, which takes the form of a multidisciplinary, written, research proposal that is related to, but not identical to, the student's dissertation research, and its oral defense in front of an exam committee assigned by the GST program (learning goals 1, analytical thinking; 2, research competence; 5, communication skills). This format is intended to ascertain an even and fair standard across the diverse program. Students often organize mock defenses for one another, an element of peer mentoring that is encouraged with pizza by the GST program.

In summary, the GST program is founded on a structured core curriculum. Other cross-cutting programs have adopted a similar strategy in order to absorb and integrate students who

come to the program from diverse backgrounds (Smith *et al.*, 2013; van de Ven *et al.*, 2015).

Transferable Skills

Original research becomes the primary focus after the comprehensive exam. Traditionally, a PhD dissertation is the work of one person, requiring accountability, yet many GST students also become part of larger research consortia that reach beyond their adviser's research group. This is particularly true for students at ORNL. Because the majority of GST students will later pursue research-related careers, the emphasis on research is certainly warranted. However, because graduate programs should allow for other professional skills to be developed, many such activities are interwoven into the curriculum, and only the major ones can be listed (learning goals: 3, teamwork and professional development; and 5, communication).

- GST students are encouraged to form a peer-learning community. For those students who actively maximize these contacts, their GST community can become the core of a multidisciplinary network that returns valuable professional contacts and companionship well beyond graduation. Besides working together through the core curriculum and challenging each other on their mock comprehensive exams, every Spring semester, all GST students join together in the weekly "colloquium." The colloquium aims to connect junior and senior students vertically. The format of the colloquium varies. For example, groups of GST students have organized a lecture series of invited speakers, including selecting and inviting speakers, preparing their visits, and hosting the speakers. This is logistically challenging and requires oversight. In another season, groups of GST students developed tutorials in their area of computational specialization and taught the tutorials to the entire class (see the Supplemental Material).
- GST students are encouraged to become stewards of their program and their discipline (Ullrich *et al.*, 2014). For example, the key committees for admissions and curriculum include one senior graduate student representative. Current students host candidates for admission to the program, and current students are paired as mentors with first-year students. Students also serve as teaching assistants for the computer programming lab. The programming lab was first designed by GST students in 2007. The lab has since been revised several times to adapt to trends in programming practice, for example, by substituting the Perl programming language with Python.
- Oral communication skills. In alternating years, the colloquium serves as a venue for students to present their progress in their dissertation research. Because GST stretches all the way from amino acids to populations, students ought to frame their project in a much broader context than they would for a presentation to their lab or their dissertation committee, or even a conference. Teaching assignments, journal clubs, and an annual poster competition also provide opportunities to develop oral communication skills. Some students are drawn to engage in outreach activities in the local education community. For example, GST students have hosted a summer camp to teach protein structural biology in UT's KidsU program to middle school students. These outreach activities are entirely voluntary.

4. Career development. GST students receive a substantial financial contribution to attend up to one conference per year. Students are also encouraged to mingle with external speakers who come to campus. The program invites GST alumni from industry and academia to our annual retreat and occasionally hosts speakers with science-related responsibilities such as head-hunting or scientific publishing to share their perspectives. The GST program serves as a conduit for information about university-wide career development workshops and also partners with other units to organize or fund such events. This is an area where the program hopes to do more in the future.

In summary, many of the programmatic elements of GST fit the mold of typical PhD programs, including other programs at UT and elsewhere. However, GST also has features that distinguish it from other programs at UT. Besides the ORNL connection, the broad spectrum of scientific topics, and the emphasis on computational training, these are 1) four GST-specific core courses; 2) a recruiting strategy targeting applicants with computational inclination; 3) financial support to attend conferences; and 4) a reduced teaching load, as low as two semesters overall.

RESULTS

The GST program has graduated PhD and MS students who have completed the GST curriculum for the past 15 years. Here, we report summary statistics to characterize the demographics of GST students, their academic background, and the career trajectories of GST alumni in light of their training on the spectrum of wet-lab and dry-lab biology. Where appropriate we make comparisons with GST's sister program, BCMB, at UT. BCMB is closely related to GST, in that most BCMB faculty members are also members of GST, and GST and BCMB students take some of the same core courses.

GST Attracts a Distinct Clientele of Students

To determine whether GST has attracted a unique population of students, we compared the majors of matriculating students between 2010 and 2016 in GST with BCMB (Table 3; see Table 1 for demographic information). First, the GST program has recruited students from a diverse range of scholarly backgrounds (Table 3), including the physical and computational sciences and bioinformatics, some of whom would not necessarily fit into the traditional departmental graduate programs at UT. Students in

BCMB have come more predominantly from the biological and chemical sciences. In recent years, GST admitted a larger fraction of international students (Table 1) who were responsible for the higher proportion of recruits who came with master's degrees.

GST Alumni Pursue Diverse Career Paths

Next, we characterized the career trajectories of PhD alumni from the GST program and compared these with career paths of alumni from the BCMB sister program. We sought to address whether the distinct structure and clientele of GST are associated with distinct outcomes. We emphasize that none of our statements should be considered as an evaluation of one outcome being qualitatively better than another. We also do not address whether GST is meeting its formal learning goals, or whether the outcomes of GST are worse or better than those of other programs.

Between 2003 and 2016, there were 77 PhDs and 22 MS graduates in GST. Of the PhDs, three out of four entered into traditional postdoctoral positions (academic and governmental labs). Most alumni remained in the United States (Table 4, see note a). A transition into a secondary position typically occurred within 4 to 8 years (Table 5).

GST alumni entered professional career paths in academia, government, or industry. The fractions in each sector were similar to those reported elsewhere (Tilghman and Rockey, 2012; University of California–San Francisco and Graduate Division, 2013), except that the fraction in government was larger (Figure 1). The trend toward governmental science was seen regardless of whether students graduated with an ORNL adviser or a UT adviser (unpublished data).

As GST PhD alumni moved on into secondary positions, at least half of those who were in academia or government switched to a different sector, often toward industry (Figure 2). International (42/76) and domestic (34/76) students followed similar career paths (unpublished data). These data support the concept of a network of branching career development pathways (Fuhrmann *et al.*, 2011). A small fraction of PhD alumni (7/39) have entered traditional academic career tracks as research-intensive principal investigators with responsibilities for training PhD students (data not tabulated). We briefly examined one major quantitative metric of graduate student success, publications. Among 77 PhD graduates from GST between 2003 and 2016, 99% coauthored at least one article and 82% published as a first author. The median number of

TABLE 3. Fields of formal training reported by new recruits to the GST and BCMB programs (2010–2016)^a

	GST			BCMB		
	Undergrad major	Master's major	Total	Undergrad major	Master's major	Total
Biological sciences (various)	17	12	29	43	13	56
Biotechnology or bioengineering	4	12	16	1	2	3
Bioinformatics	2	5	7	0	0	0
Chemistry or chemical engineering	3	0	3	6	2	8
Other	2	0	2	2	1	3
Total	28	29	57	52	18	70

^aWhen new recruits report multiple degrees or double majors, only the most recent degree or the degree most related to GST or BCMB is reported. Majors reported under "Other" include physics, computer science, food science, political science, and economics. Undergraduate majors are not reported for master's majors. Figures that are significantly elevated as compared with the other program (Fisher's exact test with $p < 0.05$) are marked in bold.

TABLE 4. GST and BCMB trainees' career paths (PhD, 2003–2016)^a

Position after PhD	GST		BCMB	
	First	Secondary	First	Secondary
Type of position				
Postdoctoral research	59		50	
No postdoctoral position	17		11	
Type of science				
Experimental biology (wet-lab)	25	9	49[†]	21[†]
Computational science (dry-lab)	30[†]	12[†]	4	1
Hybrid (wet- and dry-lab)	19[†]	8[†]	1	1
Instruction or management	0	9	6[†]	8
Other or science-unrelated	3	1	5	0
Sector				
Academia, primarily teaching	0	4	4[†]	8
Academia, primarily research	45	11	45	14
Industry, hospitals, R&D	11	17	4	7
Governmental science	18[†]	5	6	2
Other (finance, healthcare, etc.)	2	2	2	0
Geographic location				
United States	72	35	61	26
Overseas	4	4	2	5
Total	76	39	65	31

^aThis table summarizes data from GST for all except one of its 77 PhD alumni between 2003 and 2016, half of whom graduated between 2012 and 2016. Numbers of master's graduates are too small to discern trends. The secondary position is usually the current position; in cases in which the current position could not be ascertained, it is the most recent position. The secondary position is usually the second position, but for alumni who changed jobs again, it may be a later position (less than 10% of alumni). The table also summarizes data from BCMB's 65 PhD alumni; of these, 4 had incomplete information. Figures that are significantly elevated as compared with the other program (Fisher's exact test with $p < 0.05$) are marked in bold and with a dagger (†).

first-author publications was two, and the median number of all coauthorships was four. Eighty-six percent of students had published from their doctoral program before the end of the year of their defense. Because the field of research is a major predictor of the number of publications, we did not explore how the number of publications predicts subsequent career development such as employment sector or continued engagement in original research.

Of note, of the 25 MS graduates in GST, six took computational science jobs, six entered the health industry, at least four entered PhD programs; as few as three continued in wet-lab research outside of a PhD program (data not tabulated). Interestingly, among six students who completed a concurrent master's

in statistics along with their PhD or MS in GST, none accepted a traditional academic postdoc, and almost all entered straight into industry or government positions (data not tabulated).

Similar trends were found among alumni from the BCMB graduate program. About three-quarters of the PhDs entered into postdoc positions (50/61; Table 4), and the typical time for transition into a secondary position was also 4 to 8 years (Table 5). There was no significant difference between the career paths of international and domestic students, and most alumni stayed in the United States. Also similar to GST, when alumni in the academic sector transitioned into a second position, about half switched to a different sector (Figure 2). Comparing the programs, more BCMB alumni than GST alumni entered into academic teaching right after their PhD, while more GST alumni joined governmental science (significant by Fisher's two-sided test) or industry (marginally significant, $p = 0.08$; Table 4 and Figure 2).

GST Trains Students for Dry-Lab/Computational, Wet-Lab/Experimental, and Hybrid Career Paths

How did GST alumni branch out into dry-lab/computational or wet-lab/experimental career paths? How many became "hybrid" biologists? And are the respective computational and experimental emphases of the GST and BCMB programs predictive of the subsequent career paths? The dissertation research of GST's PhD alumni who graduated between 2003 and 2016 was primarily computational for 32% and hybrid for 17%, with the remaining 51% working primarily experimentally (data not tabulated). Indeed, in their first position after the PhD, a larger fraction of GST alumni than BCMB alumni assumed computational or hybrid positions, while a larger fraction of BCMB alumni took on wet-lab/experimental positions and instructional or managerial roles (Table 4). In their first position, those who worked exclusively computationally for their dissertation all remained computational, while those with hybrid dissertations tended to remain hybrid or became subsumed into fully computational roles. Among those with primarily experimental dissertations, ~33% joined laboratories with significant dry-lab components; ~10% took on computational positions, and ~25% took on hybrid responsibilities (data not tabulated). In summary, there is a clear trend among GST alumni toward computational life science. In contrast, almost all of the BCMB alumni who trained as wet-lab/experimental biologists during their PhD studies continued on the same path. The four BCMB alumni who assumed computational biology positions (Table 4) had trained in dry-labs during their PhD work (data not tabulated).

TABLE 5. Career transitions (PhD graduates between 2003 and 2016)^a

	GST				BCMB			
	Count	In first position	In later position	Percent	Count	In first position	In later position	Percent
2003–2008	21	4	17	81	20	1	19	95
2009–2012	25	8	17	68	20	9	11	55
2013–2014	15	10	5	33	7	6	1	14
2015–2016	15	15	0	0	12	12	0	0
Total	76	37	39		59	28	31	

^aAlumni who are classified as being in their first position (as of February 2017) may have received promotions or may have changed responsibilities within their organization. There were no significant differences between the two programs.

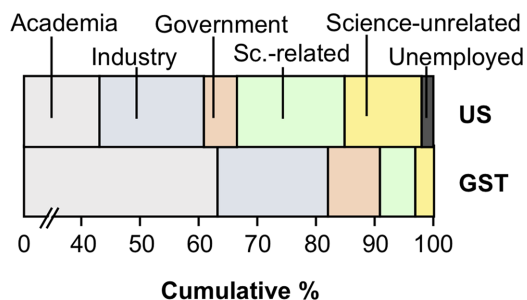


FIGURE 1. Medium-term career tracks of PhD alumni from GST 6–16 years after graduation. Of PhD alumni, 63% were in academia given the most recent data available (compare 43% nationwide), 19% in industry, 9% in government, 6% in science-related careers, and 3% in science-unrelated careers. None were known to be unemployed. Data are compared with national estimates from 2008 (Tilghman and Rockey, 2012). The GST data are from PhD graduates from 2000 to 2009.

Is it likely that a student who joined GST with a prior degree in the physical or computational sciences will become a wet-lab/experimentalist? Among 12 recent recruits who came with significant research experience in computational science or physics, none became wet-lab/experimentalists during their degree program; this includes recruits with degrees in a biological science. This even though the GST program strongly encourages our first-year students to explore wet-lab/experimental work during the lab rotations. In contrast, of 24 recent students who came grounded in wet-lab biology, 14 have

become hybrid or dry-lab researchers. Evidently, in our environment, it is much more common for wet-lab scientists to transition into dry-lab research than the other way around. Because GST students who come to the program with traditional bioscience degrees often assume dry-lab roles, we conclude that the GST program is facilitating this transition.

DISCUSSION

In summary, the GST program attracts a distinct cohort of matriculating students, it trains students for original research in wet-lab, dry-lab, and hybrid life science, and GST alumni are prepared to assume diverse career paths in a broad range of sectors. GST attracts students from traditional biosciences as well as physical science backgrounds. However, it is much more common for bioscience recruits to assume dry-lab computational projects than for physical science recruits to enter into wet-lab/experimental research. Therefore, as we strive to train hybrid biologists who have hands-on experience with the structure, provenance, and limitations of experimental data, our data suggest that these scientists will more likely come from experimental biology rather than the computational or physical sciences. This finding might also suggest that there is a niche for so-inclined physical scientists to warm up to wet-lab experimental research if they bring a unique perspective to their research that adds a competitive edge over the admittedly large pool of wet-lab experimentalists with traditional training.

Our conclusions were possible because the GST program carefully tracks the career trajectories of its alumni as far as publicly available records and personal contacts allow. Of course, we respect requests for privacy by our alumni. In keeping with directives from the training and workforce development leadership at the NIH (Singh *et al.*, 2015) and calls for transparency from the community of trainees (McDowell *et al.*, 2014; Polka *et al.*, 2015), we publish the whereabouts of our alumni on the GST Web page. We do this to communicate realistic expectations among our potential applicants and stakeholders, inspire current students, and promote networking between students and alumni. As of today, very few graduate programs make their outcome data publicly available (Gibbs, 2016).

The dearth of data sets, not to speak of analytical studies, on outcomes in graduate education is disconcerting, given the competitive and global job market, given that the pursuit of happiness of individuals is at stake, and given that billions of dollars are being invested in graduate education across the United States. Agencies that fund graduate training wish to support evidence-based innovative approaches (NSF, 2014a; Gammie *et al.*, 2017; National Academies of Sciences, Engineering, and Medicine, 2017), and there has been a call for a framework for collecting standardized, numerical outcome data that will lend themselves to aggregation into (meta-)analyses with sufficient statistical power (Gibbs and Marsteller, 2016).

Much of the effort on analyzing the quality of graduate programs has traditionally focused on faculty research metrics with little if any consideration given to the preparation of the incoming students or their career outcomes. Instead, it may be fruitful to address what specific factors among individual students contribute to specific career outcomes for individual students. In our analysis here, we considered prior training of matriculated

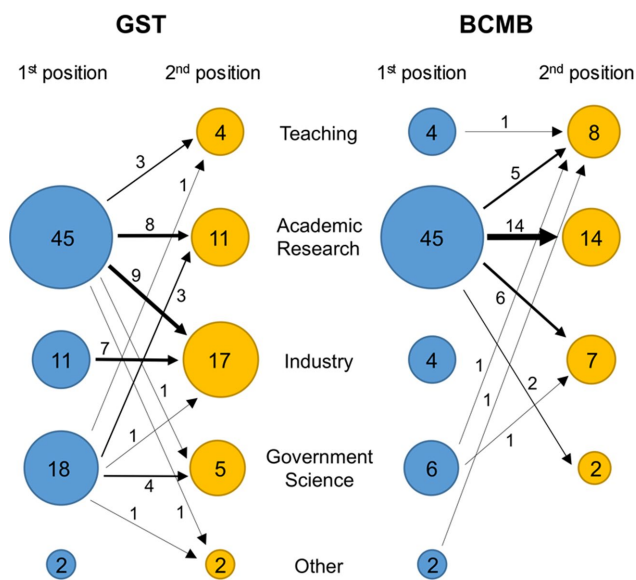


FIGURE 2. Career transitions for alumni from the GST program (left) and BCMB (right). For each program, the left column shows the sector of the graduate’s first position after PhD and the right column shows a subsequent position, usually the second position. The numbers reflect numbers of individuals, not percentages; note that about half of the alumni have not yet moved into a second position. Transitions from one sector to another are depicted by arrows; the flux is indicated by the number of individuals and line thickness.

students (e.g., prior major), we considered what the students experienced in their PhD programs (GST or BCMB), and we considered the career outcomes of the students. And the analysis did yield statistically significant results, despite relatively small numbers of trainees.

To go further and identify additional predictors of graduate training outcomes, whether interventions or contextual factors such as demographics, will be difficult. Graduate programs are usually small. Moreover, each of their numerous training elements and the interactions among those elements are bound to contribute no more than a small portion to outcomes, and randomized controlled trials are not practical. Therefore the necessary statistical power to identify factors for success may require that data from multiple programs be aggregated, a practice common in other educational settings (Byars-Winston and Fouad, 2008; Gibbs *et al.*, 2014; St. Clair *et al.*, 2017). Adding variables, such as geographic location, differences in curricula, and admissions criteria, is not necessarily a disadvantage, because the effect of these variables can then be explored. In addition, we point out that similar graduate programs already recruit from pools of applicants that are substantially overlapping, and most programs prepare their graduates for a single, increasingly global, job market. Therefore, we suggest that aggregating data across multiple programs would be useful to determine what factors among students and programs predict what outcomes. For example, when it comes to training scientists in the expanding field of computational biology, is outcome determined by the inclination of the recruits before entering a program? Our data suggest “yes, to some degree.” Is it driven by their experience while in the program? Transitions by wet-lab biologists into dry-lab research suggest “yes.” Or is it driven solely by the pull of the job market, which has demanded and absorbed computationally enabled biologists? This is not likely; if that were the case, BCMB and GST alumni would get drawn into computational biology jobs to an equal degree. More comprehensive information of this kind would help so-inclined graduate programs to innovate with more confidence and would assist future applicants in finding the program that is right for them.

Our study has a number of limitations. First, while most of our outcome measures are fairly objective, the definition of what constitutes a hybrid researcher is ad hoc, may change as the field develops, and also involves self-assessment on the part of students or alumni. Yet we feel that this categorization was useful. Second, no attempts were made to evaluate which experiences, program elements, or accomplishments were predictive of which outcomes. Third, our study did not present any perceptions of students, faculty, or alumni, regarding what factors, intrinsic or extrinsic to the GST program, led to their reported career outcomes. A mixed-methods study that couples data on student experiences and outcomes could help address this. Such a study might also benefit from a more robust theoretical framework from social cognitive career theory or expectancy value theory, as espoused by several recent studies of life science education (Byars-Winston *et al.*, 2010; Bierer *et al.*, 2015; Gibbs *et al.*, 2015; St. Clair *et al.*, 2017). These theories posit that motivations, career interests, and eventual career goals are shaped by a person’s self-efficacy (perception of one’s own competency), outcome expectations, and values, as well as by contextual factors, such as demographics (Lent *et al.*, 1994; Eccles

and Wigfield, 2002). Self-efficacy, outcome expectations, and values can be measured using real-time survey instruments. The benefit of measuring these constructs is that they aggregate the effect of innumerable environmental and innate factors from a person’s life history into a few measurable parameters that have been shown to predict her or his motivation and goals (Byars-Winston and Fouad, 2008; Gibbs and Griffin, 2013; Bierer *et al.*, 2015).

ACKNOWLEDGMENTS

We thank the former and current students in the GST program for sharing personal information without which an assessment of the program would be impossible. We thank the GST faculty for all their enthusiastic grassroots support for the program. We thank former GST directors Jeff Becker and Cynthia Peterson for their vision in creating and growing the program and launching it on its subsequent path. We thank the College of Arts and Sciences for financial support. Research in the von Arnim lab is supported by NSF-MCB-1546402 and NSF-IOS-1456988. The sponsor of the study had no role in the design or conduct of the study or in the decision to publish. We thank the NSF and the NIH for their independent yet largely coherent perspectives on the future of postgraduate training in computationally enabled life science.

REFERENCES

- Alberts, B. (2012). Creativity at the interface. *Science*, 336, 131.
- Alon, U. (2009). How to choose a good scientific problem. *Molecular Cell*, 35(6), 726–728. doi: 10.1016/j.molcel.2009.09.013
- Benfey, P. (2013). Philip Benfey. Interview. *Current Biology*, 23(19), R857–R858.
- Bierer, S. B., Prayson, R. A., & Dannefer, E. F. (2015). Association of research self-efficacy with medical student career interests, specialization, and scholarship: A case study. *Advances in Health Sciences Education Theory and Practice*, 20(2), 339–354. doi: 10.1007/s10459-014-9531-7
- Bourne, H. R. (2013). A fair deal for PhD students. *eLife*, 2, e01139.
- Brokaw, J. J., & O’Loughlin, V. D. (2015). Implementation of an education-focused PhD program in anatomy and cell biology at Indiana University: Lessons learned and future challenges. *Anatomical Sciences Education*, 8(3), 258–265. doi: 10.1002/ase.1482
- Byars-Winston, A., Estrada, Y., Howard, C., Davis, D., & Zalapa, J. (2010). Influence of social cognitive and ethnic variables on academic goals of underrepresented students in science and engineering: A multiple-groups analysis. *Journal of Counseling Psychology*, 57(2), 205–218. doi: 10.1037/a0018608
- Byars-Winston, A., & Fouad, N. A. (2008). Math and science social cognitive variables in college students contributions of contextual factors in predicting goals. *Journal of Career Assessment*, 16, 425–440.
- Daniels, R. J. (2015). A generation at risk: Young investigators and the future of the biomedical workforce. *Proceedings of the National Academy of Sciences USA*, 112(2), 313–318. doi: 10.1073/pnas.1418761112
- Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology*, 53, 109–132. doi: 10.1146/annurev.psych.53.100901.135153
- Fuhrmann, C. N., Halme, D. G., O’Sullivan, P. S., & Lindstaedt, B. (2011). Improving graduate education to support a branching career pipeline: Recommendations based on a survey of doctoral students in the basic biomedical sciences. *CBE—Life Sciences Education*, 10(3), 239–249. doi: 10.1187/cbe.11-02-0013
- Gammie, A., Gibbs, K. D., & Singh, S. (2017). Early Notice: New NIGMS Institutional Predoctoral Training Grant Funding Opportunity Announcement. Retrieved January 3, 2017, from <https://loop.nigms.nih.gov/2017/03/early-notice-new-nigms-institutional-predoctoral-training-grant-funding-opportunity-announcement>.

- Gibbs, K. D. (2016). Analysis of NIGMS Request for Information on Strategies for Modernizing Biomedical Graduate Education. Retrieved January 3, 2017, from NIGMS Feedback Loop Blog. www.nigms.nih.gov/training/Documents/graduate-education-rfi-analysis.pdf
- Gibbs, K. D., & Griffin, K. A. (2013). What do I want to be with my PhD? The roles of personal values and structural dynamics in shaping the career interests of recent biomedical science PhD graduates. *CBE—Life Sciences Education*, 12(4), 711–723. doi: 10.1187/cbe.13-02-0021
- Gibbs, K. D., & Marsteller, P. (2016). Broadening participation in the life sciences: Current landscape and future directions. *CBE—Life Sciences Education*, 15(3), ed1.
- Gibbs, K. D., McGready, J., Bennett, J. C., & Griffin, K. (2014). Biomedical science Ph.D. career interest patterns by race/ethnicity and gender. *PLoS ONE*, 9(12), e114736. doi: 10.1371/journal.pone.0114736
- Gibbs, K. D., McGready, J., & Griffin, K. (2015). Career development among American biomedical postdocs. *CBE—Life Sciences Education*, 14(4), ar44. doi: 10.1187/cbe.15-03-0075
- Griffiths, P. A., Adams, R. M., Alberts, B. M., Bement, A. L., Blout, E. R., Browder, F. E. ... McCray, L. E. (1995). *Reshaping the graduate education of scientists and engineers*, National Academies Press.
- Hogan, A., & Roberts, B. (2015, December). Occupational employment projections to 2024. *Monthly Labor Reviews*, pp. 1–25.
- Johnson, C. Y. (2014, October 5). Glut of postdoc researchers stirs a quiet crisis in science. *Boston Globe*, p. A-1.
- Kennedy, D., Austin, J., Urquhart, K., & Taylor, C. (2004). Supply without demand. *Science*, 303, 1105.
- Kuehnle, K., Winkler, D. T., & Meier-Abt, P. J. (2009). Swiss national MD-PhD program: An outcome analysis. *Swiss Medical Weekly*, 139(37–38), 540–546. doi: sMw-12790
- Lacey, T. A., & Wright, B. (2009). Occupational employment projections to 2018. *Monthly Labor Reports*, 132, 82–123.
- Layton, R. L., Brandt, P. D., Freeman, A. M., Harrell, J. R., Hall, J. D., & Sinche, M. (2016). Diversity exiting the academy: Influential factors for the career choice of well-represented and underrepresented minority scientists. *CBE—Life Sciences Education*, 15(3), ar41. doi: 10.1187/cbe.16-01-0066
- Lent, R. W., Brown, S. D., & Hackett, G. (1994). Toward a unifying social cognitive theory of career and academic interest, choice, and performance. *Journal of Vocational Behavior*, 45, 79–122.
- Leshner, A. (2015). Editorial: Rethinking graduate education. *Science*, 349, 349.
- McDowell, G. S., Gunsalus, K. T., MacKellar, D. C., Mazzilli, S. A., Pai, V. P., Goodwin, P. R., ... Polka, J. K. (2014). Shaping the future of research: A perspective from junior scientists. *F1000Res*, 3, 291. doi: 10.12688/f1000research.5878.2
- National Academies of Sciences, Engineering, and Medicine. (2017). Revitalizing Graduate STEM Education for the 21st Century. Retrieved January 3, 2017, from <https://sites.nationalacademies.org/PGA/bhew/graded/index.htm>
- National Research Council. (2011). *Research training in the biomedical, behavioral, and clinical research sciences*, Washington, DC: National Academies Press. Retrieved January 3, 2017, from www.ncbi.nlm.nih.gov/books/NBK56994/pdf/Bookshelf_NBK56994.pdf
- National Science Board. (2014). Science and Engineering Indicators 2014. Retrieved January 3, 2017, from www.nsf.gov/statistics/seind14/content/etc/nsb1401.pdf
- National Science Foundation (NSF). (2014a). National Science Foundation Research Traineeship (NRT) Program: Program Solicitation. NSF 14-548. Retrieved January 3, 2017, from www.nsf.gov/pubs/2014/nsf14548/nsf14548.htm
- NSF. (2014b). Survey of Earned Doctorates Doctorate Recipients from U.S. Universities: 2013. NSF 15-304. Retrieved January 3, 2017, from <https://www.nsf.gov/statistics/sed/2013/>
- NSF. (2014c). Survey of Graduate Students and Postdoctorates in Science and Engineering, Fall 2012. Retrieved January 3, 2017, from <https://ncesdata.nsf.gov/gradpostdoc/2012/index.html>
- Noble, D. B., Mochrie, S. G., O'Hern, C. S., Pollard, T. D., & Regan, L. (2016). Promoting Convergence: The Integrated Graduate Program in Physical and Engineering Biology at Yale University, a New Model for Graduate Education. *Biochemistry and Molecular Biology Education*, 44(6), 537–549. doi: 10.1002/bmb.20977
- Petsko, G. (2003). Q & A. *Current Biology*, 13(20), R787–R788.
- Polka, J. K., Krukenberg, K. A., & McDowell, G. S. (2015). A call for transparency in tracking student and postdoc career outcomes. *Molecular Biology of the Cell*, 26(8), 1413–1415. doi: 10.1091/mbc.E14-10-1432
- Saueremann, H., & Roach, M. (2012). Science PhD career preferences: Levels, changes, and advisor encouragement. *PLoS ONE*, 7(5), e36307.
- Singh, S., Gammie, A., & Lorsch, J. (2015). Catalyzing the Modernization of Graduate Education. NIGMS Feedback Loop blog. Retrieved January 3, 2017, from <https://loop.nigms.nih.gov/2015/11/catalyzing-the-modernization-of-graduate-education/>
- Smith, C. L., Jarrett, M., & Bierer, S. B. (2013). Integrating clinical medicine into biomedical graduate education to promote translational research: Strategies from two new PhD programs. *Academic Medicine*, 88(1), 137–143. doi: 10.1097/ACM.0b013e31827644ef
- St. Clair, R., Hutto, T., MacBeth, C., Newstetter, W., McCarty, N. A., & Melkers, J. (2017). The "new normal": Adapting doctoral trainee career preparation for broad career paths in science. *PLoS ONE*, 12(5), e0177035. doi: 10.1371/journal.pone.0177035
- Teitelbaum, M. S. (2014, March 19). The myth of the science and engineering shortage. *The Atlantic*. Retrieved January 3, 2017, from www.theatlantic.com/education/archive/2014/03/the-myth-of-the-science-and-engineering-shortage/284359/
- Tilghman, S., & Rockey, S. (2012, June 14 2012). National Institutes of Health: Biomedical Research Workforce Working Group Report. Retrieved January 3, 2017, from https://acd.od.nih.gov/biomedical_research_wgreport.pdf
- Ullrich, L., Dumanis, S. B., Evans, T. M., Jeannotte, A. M., Leonard, C., Rozzi, S. J., ... Forcelli, P. A. (2014). From student to steward: The Interdisciplinary Program in Neuroscience at Georgetown University as a case study in professional development during doctoral training. *Medical Education Online*, 19, 22623. doi: 10.3402/meo.v19.22623
- University of California—San Francisco and Graduate Division. (2013). Aggregate Data for Graduate Division Programs. Retrieved December 18, 2015, from <https://graduate.ucsf.edu/aggregate-data>
- van de Ven, A. L., Shann, M. H., & Sridhar, S. (2015). Essential components of a successful doctoral program in nanomedicine. *International Journal of Nanomedicine*, 10, 23–30. doi: 10.2147/ijn.s69144
- Woolston, C. (2015). Graduate survey: Uncertain futures. *Nature*, 526(7574), 597–600.
- Yamamoto, K. (2014). Time to rethink graduate and postdoc education. *iBiomagazine*, (11). Retrieved January 3, 2017, from www.ibiology.org/ibiomagazine/issue-11/keith-yamamoto-time-to-rethink-graduate-and-postdoc-education.html