NATIONAL HUMAN GENOME RESEARCH INSTITUTE/ NIH/DHHS

SUMMARY

FIFTH WORKSHOP

OF THE

TRAINING COORDINATORS

22-23 April 2013
5th floor conference room
5625 Fishers Lane
Rockville, MD 20852

TABLE OF CONTENTS

FIFTH TRAINING COORDINATORS' WORKSHOP

SUMMARY

FIFTH TRAINING COORDINATORS' WORKSHOP

I. INTRODUCTION

Bettie Graham opened the meeting by welcoming the participants and introducing the two new NHGRI Training Program Directors, Heather Junkins and Tina Gatlin. Bettie reiterated the main purpose of meeting: (1) to review progress toward meeting the goal of increasing the number of underrepresented minorities (URMs) in genomic sciences; (2) to review the quality of the data submitted to REDCap; and (3) to provide a mini-grant writing and grants management workshop. The underlying theme was to stress to the training coordinators' the importance of entering data into REDCap.   REDCap is the tracking system used by the Washington University, St. Louis data analysis and coordinating center (DACC), to track and evaluate NHGRI's DAP and research training programs to provide reports that would demonstrate NHGRI's progress in meeting its goal of increasing the number of URMs participating in genomics research, and increasing the number of researchers staying in genomic sciences. Other activities that were part of the meeting were: (1) progress reports by a representative from each DAP and T32 program; (2) two report outs by the DACC; and (3) and a general discussion. The Agenda (Appendix I) and Roster (Appendix II) are attached.

II. PROGRESS REPORTS

In advance of the meeting, Heather Junkins and the DACC provided the training centers PowerPoint presentation templates in an effort to standardize the 10 minute presentations from each of the groups. Of the ten template slides, six consisted of tables to be filled out in an effort to ensure uniform data reporting amongst the groups. Prior to the meeting, the DACC encouraged the PIs to update their data in REDCap. The DACC then "froze" the data in REDCap on March 29, 2013, and then provided pre-filled tables for each of the centers as a courtesy to aid in gathering the data for their presentations.   Data slides included: 1) a program description outlining specific programs with duration, education level, number of trainees, and IRB status; 2) training components; 3) demographics; 4) URM sample sizes; and 5) metrics of success. The centers were also charged to include descriptive information on individual factors contributing to the success of their programs and any challenges they have faced, and given an extra slide to personalize their report.

a. **T32 Institutional Training Grants**

There are twelve institutional T32 training centers. Instead of detailing the REDCap data report outs from each of the center presentations, aggregate data (provided by the DACC) is presented in Section III. Briefly, notable aspects from each center are highlighted below:

University of California, Berkeley – presented by Jasper Rine (co-PI)

This T32 program is in its 13th year and is mostly focused on pre-docs. Some factors noted as contributing to their success is the quality/record of the training faculty, new faculty hires to training grant, and the highly collaborating faculty. Outreach efforts include a GED program at San Quentin State Prison; tutors for local high schools; and attendance at ABRCMS & SACNAS. They have struggled with recruiting URMs. Stated challenges include competition for RO1 funding, California state support decrease, and faculty poaching from private institutions.

Data entry status: Baseline data entered on 122 individuals, with 54 individuals (44% of baseline records) having follow-up data (total target sample size unknown).

University of California, Los Angeles – presented by Jeanette Papp (Training Coordinator)

The UCLA T32 is in its 15th year as a full-time, all year round pre-doc program. Eight trainees are supported each year. A success is recruitment of URMs (representing 22% of trainees) and good retention (only one has left the program before the second year). A major challenge is the "fierce competition for a small pool of talented URM PhD candidates". This is a recurring theme for many coordinators which should be addressed.

Data entry status: Baseline data entered on 46 individuals, with 42 individuals (91% of baseline records) having follow-up data (total target sample size unknown).

Harvard University – presented by Katherine Flannery (Program Manager)

The BIG (Bioinformatics Integrative Genomics) program has existed for 10 years and funds ~ 5 pre-docs per year. The first 9 years, MIT was the home institution, but as of September 2012, the program moved to Harvard. The PIs are still Isaac Kohane and Peter Park. Because BIG only recently moved the program to Harvard and because they are still under review by their IRB, Ms. Flannery says they have not been able to collect any REDCap data and thus had nothing to share in the presentation. A positive aspect to the program is a highly collaborative faculty.

Data entry status: Baseline data entered on 100% of individuals (N = 42), with 1 individual (2% of baseline records) having follow-up data.

Princeton University – Alison Gammie (Director of Diversity Programs)

Princeton's T32 pre-doc program has been funded for 9 years. Good success with URMs in the program (9/48; 19%). Two goals are to: 1) increase URMs (Diversity program started in 2007 and data demonstrating success); and 2) increase success of students. They have a significant number of applicants from their Princeton summer programs, which has been an important recruiting tool. The program is able to leverage funds from the Office of the Provost. Students start with an intensive summer course to develop skills and help with adapting to research culture. Major issues: undergraduate training is inadequate; culture shock - feeling alienation; having to redefine success and sense of self-worth; severe loss of confidence; sometimes depression. Alison is "in there all day every day" to keep an eye out; will meet once a week with struggling students.

Data entry status: Baseline data entered on 48 individuals, with 2 individuals (4% of baseline records) having follow-up data (total target sample size unknown).

University of Michigan – Dawn Keene (Program support and administration)

Michael Boehnke's T32 is in its 18th year, but only baseline data has been entered in REDCap. They currently support 13 trainees. The program supported has 3 URMs in 2013, and the program has worked hard to recruit and retain URM students by developing close relationships with HBCUs and the Mathematical and Theoretical Biology summer program at Arizona State University (8/77; 10% URMs over course of program). Recruitment/retention activities include: visiting and building relationships with 3 HBCUs; early entry to orient students to program; tutoring; extended mentoring; focus groups; team building; flexibility; review of all application materials by the Program Director who also sits on the admissions committee.

Data entry status: Baseline data entered on 77 individuals, with 0% of baseline individuals having follow-up data (total target sample size unknown).

University of Washington – Elena Hernandez (Training Coordinator)

This is another long-standing T32 program in its 18th year. Funding is for 14-18 pre-docs and 7 post-docs per year. 35/349 (12%) trainees are URM which is 2% above their target. Factors contributing to success include "pre-recruiting via their summer research program and their University of New Mexico weekend visit. Challenge noted is that while URM students are more prepared for study in Genomics at the doctorate level, URM recruiting at the graduate level is intense at the top institutions.

Data entry status: Baseline data entered on 184 individuals, with 4% of baseline individuals (N = 7) having follow-up data (total target sample size).

University of Wisconsin-Madison – Louise Pape (Training Coordinator)

The Wisconsin T32 is in its 10th year, and funds 10 pre-docs and 4 post-docs per year. URMs are 15% (9/59) of total trainees. Some notable factors contributing to success include: weekly interactive meeting/seminar with trainees, Program Director David Schwartz and other faculty; cross-disciplinary research collaborations; and participation at CSHL Biology of Genomes meeting each year. Challenges highlighted include: difficulties getting written consents from former trainees; and increasing the diversity of URMs particularly for post-docs (successful with Hispanics, not so for Blacks and Native American); financial pressures on trainees; and difficult job market particularly for past trainees seeking faculty positions.

Data entry status: Baseline data entered on 100% of individuals (N = 59), with 53% of individuals (N = 31) having follow-up data.

Yale University – Jenny Frederick (Training Director/Coordinator)

Since 2001, they have had 55 trainees (approximately twice as many graduate students as post-docs), with 4 being URMs (7%). Many minority recruitment efforts, but few come to Yale. Challenges include: support for first year undecided students; connecting students across a range of departments and programs and across 3 campuses; defining success beyond the traditional academic career metrics. Professional development of trainees places emphasis on scientific communication skills. Little follow-up data have been entered into REDCap.

Data entry status: Baseline data entered on 100% of individuals (N = 55), with 2% of baseline individuals (N = 1) having follow-up data.

Massachusetts Institute of Technology – Jeanne Darling (Training Coordinator)

This is a small and new T32 with David Gifford as the PI. Their PhD program in computational genetics is in its 4th year and supports 4 pre-docs per year. Of their 11 total trainees to date, they have 1 URM and 2 women. Strengths of their program include a large number of collaborative projects between students and local experts in genetics and genomics, and proximity to excellent resources in genomics and genomic medicine.

Data entry status: No data entry to date.

University of Pennsylvania; Data entry status: Baseline data entered on 94% of individuals (N = 44), with 0% of baseline individuals having follow-up data.

Washington University in St. Louis; Data entry status: Baseline data entered on 100% of individuals (N = 46), with 20% of individuals (N = 9) having follow-up data.

Stanford University; Data entry status: Baseline data entered on 192 individuals, with 192 individuals (100% of baseline records) having follow-up data (total target sample size unknown).

The above three training programs did not have a representative at the meeting. Treva Rice gave overview presentations of their data entry into REDCap, and additional information conveyed by the PIs/coordinators.

b. DAP PROGRESS REPORTS

There are eleven diversity action plan programs across the various large-scale sequencing centers, database projects and CEGS grants. Instead of detailing the REDCap data report outs from each of the center presentations, aggregate data (provided by the DACC) is presented in Section III. Briefly, notable aspects from each center are highlighted below:

Baylor College of Medicine – presented by Debra Murray (Senior Investigator)

This summer course is in its tenth year and supports a combination of undergraduates and post-baccalaureates. This course is designed to provide the necessary skills and background training in bioinformatics to bring students up to speed before the start of the academic school year. Factors contributing to the success of the program include having experienced PIs, strong relationship with undergraduate faculty, and the opportunity for students to transition to a post-baccalaureate position. Challenges include the difficulty with engaging young women to enter the field of bioinformatics and helping new students navigate the educational system. The program has been able to keep in contact with alumni and keep them engaged with quarterly meetings that focus on networking and professional development.

Baseline data entered on 116 individuals, with 97% of baseline individuals (N = 113) having follow-up data (total target sample size is unknown).

The Broad Institute – presented by Eboney Smith (Program Coordinator)

This program in its eleventh year and has provided support to students at five different career levels, although one (the medical fellows program) has been discontinued. The program currently supports students through summer and academic year programs at 4

educational levels; this includes K-12, undergraduates, post-docs, and faculty. The K-12 program introduces high school seniors to science and math. Factors contributing to the success of the program include commitment by Broad Institute staff and leadership, scientific communication course, mentor orientation, and strong relationships with other Boston-area programs. Challenges include tracking students, recouping overhead costs not covered by the grant, balancing time in and out of the lab, finding strong post-doc candidates, and identifying appropriate measures of success for faculty at non-research intensive colleges and universities.

Baseline data entered on 187 individuals, with 76% of baseline individuals (N = 142) having follow-up data (total target sample size is unknown).

Washington University at St. Louis – presented by Dawayne Whittington (Consultant)

This program is in its seventh year and supports undergraduates and post-baccalaureates for an eight-week summer program. The success of the program can be attributed to several factors; emphasis on science literacy, commitment to formally collecting feedback from students and mentors, strong research placements for students and improved student confidence as a result of the program. The program has difficulty with tracking students after they have left the program. All of the programs are dealing with this challenge.

Baseline data entered on 82 individuals, with 52% of baseline individuals (N = 43) having follow-up data (total target sample size is unknown).

University of California, Santa Cruz – presented by Zia Isola (Program Coordinator)

This program is in its tenth year and supports undergraduate and graduate students with supplemental training. The training includes mentoring, giving scientific presentations and two-minute elevator talks about their work, reading scientific literature, grant writing, and writing an annotated bibliography. There are several factors that contribute to the success of the program; good relationship with school admission committee, institutional support from UCSC, professional development workshops, and good faculty mentors that support the program. Some of the challenges the program faces are the general lack of URM students applying to computational fields, admissions committees' misunderstandings regarding California Prop. 209, proximity to Stanford and Berkeley, attraction of industry jobs in the region, and the location/design of the UCSC campus.

Baseline data entered on 100% of individuals (N = 62), with 3% of baseline individuals (N = 2) having follow-up data.

Harvard University/University of New Mexico – presented by Paul Szauter (Program Coordinator)

This program is in its fourth year and supports undergraduate students and post-baccalaureates. The training includes mentoring, skills training, networking and didactic training in biology and genetics. Challenges include following-up with trainees, dealing with cuts on research grants that support trainees, and the general use of GRE scores for graduate school entrance. These standardized tests are not strong indicators of success and this can put URM students at an unfair disadvantage.

Baseline data entered on 24 individuals, with 12% of baseline individuals (N = 3) having follow-up data (total target sample size is unknown).

The Jackson Laboratory – presented by Joanne Berghout (Program Coordinator)

This program is in its tenth year and supports high school students and undergraduates. These students work on independent projects, and training includes didactic training, mentoring, skills training, lab time for programming and data curation, networking, and a final presentation on their project. Challenges faced by the program include the fact that the JAX is not a degree-granting institution (no brand recognition), remoteness of the lab's location, and database projects may not appeal to students seeking laboratory research experience (recruitment issue).

Baseline data entered on 25 individuals, with 80% of baseline individuals (N = 20) having follow-up data (total target sample size is unknown).

Dana Farber Cancer Institute – presented by Karen Burns White (Program Coordinator)

This program is in its fifth year and supports students in high school, undergraduates and post-baccalaureates. Training is similar to other programs; didactic training, mentoring, skills training, with an expanded lab experience (computing/clinical setting/mass spec/wet lab). Factors contributing to the program success include; post-baccalaureate students complete a program work plan that focuses on academic and career planning (e.g. coursework, GRE/MCAT prep, presentation, publications). The K-12 trainees and summer undergraduates have all prepared abstracts and presented their work. Challenges facing the program include; identifying funding/resources to meet the demands of the program and evaluating the mentoring process to make sure trainees' needs are being met.

Baseline data entered on 22 individuals, with 59% of baseline individuals (N = 13) having follow-up data (total target sample size is unknown).

Harvard University – presented by Alexander Hernandez-Siegel (Program Coordinator)

This program is in its fifth year and provides training to undergraduate students and post-docs. Program elements include didactic training, mentorship, skills training and networking. Factors contributing to the success of the program include; combination of mentoring and independent work builds students' confidence, heavy recruitment in undergrad programs has yielded more URMs and high profile labs help attract the best students. The program has faced some challenges; many students are interested in pursuing MD or MPH degrees versus PhD or MD/PhD programs. Students have varying degrees of preparation in the biology and math areas, the cost of travel/living in the Boston area, and few URMs hold PhDs in related disciplines.

Baseline data entered on 30 individuals, with 10% of baseline individuals (N = 3) having follow-up data (total target sample size is unknown).

Johns Hopkins University – presented by Vicky Schneider (Program Coordinator)

This program is in its ninth year and provides training to high school students. Training includes mentorship, didactic training, laboratory experience (wet), and skills training in scientific writing, college application preparation, and career planning. The program is successful because trainees receive personalized attention from the program coordinator. The program has a long history and this has enabled long-term relationships to develop between student cohorts and alumni. Challenges include; engaging new laboratories to take on students, identifying students who will be mature enough for a lab experience, and collecting follow-up data once students graduate.

Baseline data entered on 100% of individuals (N = 79), with 80% of baseline individuals (N = 63) having follow-up data.

University of Southern California – presented by Steve Finkel (Program Coordinator)

The program is in its fifth year and supports undergraduate and graduate students. The program elements include didactic training, mentorship, skills training and networking. Program success can be attributed to significant faculty participation as mentors, informing students about coursework options, STEM careers and how to navigate transitions in an academic setting. The program also believes in treating students like grad students and this helps build confidence through independence.

Baseline data entered on 135 individuals, with 100% of baseline individuals (N = 135) having follow-up data (total target sample size is unknown).

University of Washington – presented by Elena Hernandez (Program Coordinator)

This program is in its eleventh year and supports high school students and undergraduates. The program is successful because of strong relationships with fellow participants and faculty, support and advocacy from the university, mentorships help create strong identities and development as a scientist and involvement with SACNAS. Challenges faced by the program include; reduced level of overall funding, finding faculty mentors for incoming freshmen, varying degrees of academic preparation for math and science, and finding financial aid for students.

Baseline data on 236 individuals, with 75% of baseline individuals (N = 177) having follow-up data (total target sample size is unknown).

III. REDCAP DATA SUMMARY BY THE DACC

All four members of the Washington University-based DACC were present at the meeting:
Treva Rice (PI), Karen Laseter (Data Manager), Donna Jeffe (Co-I) and Jeanne Cashman (Project Manager). Treva Rice gave two separate presentations summarizing T32 and DAP REDCap data.

a. T32 Data Summary

Treva started her presentation by emphasizing the importance of data quality. Data can be considered to be "high quality" if it is "fit for their intended uses".   Quality management can be carried out by use of the [Juran Trilogy](#), which consists of three phases: quality planning, quality control and quality improvement. Most organizations focus on quality control with little or no emphasis on the other two. To sustain and improve quality, all three phases of a management process should be clearly defined and robust.

The DACC has compiled aggregate REDCap data from the twelve T32 centers. Table 1 shows aggregate demographics. There are 915 baseline records across the programs, with 339 follow-up records (maximum of 1 follow-up visit per person). Thus, there are follow-ups on ~ 1/3 of trainees. The goal is to have follow-up data on EVERY trainee. There are slightly more males than females (M:F, 62:38).

Table 1. T32 Aggregate Demographics

Sample Sizes	Program	Miss	Pre-Doc	Post-Doc	Combined	Total
Target Sample Size						
Total Baseline		5	634	199	77	915
Follow-up Individuals		2	261	76	0	339
Refused Follow-up Consent			4	10		14
Female			232	75	42	349
Male		5	398	122	35	560
Missing Sex			4	2		6

Table 2 shows aggregate URM sample sizes. 7% either refused or left the race question blank. Race definitions are based on OMB regulations. 12% of trainees classify themselves as URMs, with about half of those identifying as Hispanic.

Table 2. T32 Aggregate URM Sample Sizes

	Missing	Pre-Doc	Post-Doc	Combined	Total	
Undefined	0	13	1	0	14	1%
Refused	0	37	14	2	53	6%
NOT Underrepresented Minority Groups						
White	4	397	135	57	593	
Asian	1	94	32	10	137	
>1 race		8	1	2	11	
Subtotal	5	499	168	69	741	81%
Underrepresented Minority Groups						
Black		24	2	3	29	
Am Ind		6	2		8	
Pac Isl		3	2		5	
Hispanic		39	10	2	51	
>1 race		13		1	14	
Subtotal	0	85	16	6	107	12%
Total	5	634	199	77	915	

Treva presented several tables of aggregate metrics of success. It is difficult to draw conclusions because of the data that are still missing. For example, entered data show that 65% of trainees remain in STEM, but the other nearly 35% remain unknown due to MISSING data. In fact, only about 2.6% definitely did NOT remain in STEM.

b. DAP Data Summary

Treva Rice gave the presentation on the REDCap data reported from the DAP programs. The programs have a responsibility to collect and enter data into REDCap and the DACC has the responsibility for monitoring data quality and implement improvements to the database as needed. Race and ethnicity are collected via two questions from the OMB regulations. Both questions need to be answered in order for the data to be complete; this includes a question about (1) Identifying self as Hispanic or Latino; (2) Selecting one or more races as they apply to the individual. Treva reviewed the different race and ethnicity categories and a summary is provided below:

American Indian and Alaska Native
    (1) People having origins in any of the original peoples of North and South America, including Central America, and who maintain tribal affiliation or community attachment. Should be accompanied by request for tribal affiliation when possible

Native Hawaiian or Other Pacific Islander
    (2) A person having origins in any of the original peoples of Hawaii, Guam, Samoa or other Pacific Islands Carolinian, Fijian, Kosraean, Melanesian, Micronesian, Northern Mariana Islander, Palauan, Papua New Guinean, Ponapean (Pohnpelan), Polynesian, Solomon Islander, Tahitian, Tarawa Islander, Tokelauan, Tongan, Trukese (Chuukese), Yapese.

Asian
    (3) A person having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent including Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, Philippine Islands, Thailand, Vietnam Afghanistan, Armenia, Azerbaijan, Bangladesh, Bhutan, Burma, East Timor, Georgia, Indonesia, Kazakhstan, Kyrgyzstan, Laos, Macau, Maldives, Mongolia, Nepal, Pakistan, Taiwan, Sri Lanka, Tajikistan, Turkmenistan, Uzbekistan.

White
    (4) A person having origins in any of the original peoples of Europe, the Middle East, or North Africa Middle East: Egypt, Turkey, Iran, Iraq, Saudi Arabia, Yemen, Syria, Israel, Jordan, United Arab Emirates, Lebanon, Oman, Kuwait, Qatar, Bahrain, Cyprus
    (5) North Africa: Algeria, Egypt, Libya, Morocco, Tunisia, Western Sahara (Sudan?)

Black or African American
    (6) A person having origins in any of the black racial groups of Africa

Table 3 summarizes the types of errors in classifying race and clarifies how these individuals would be classified under OMB regulations.

Table 3. Types of errors in classifying race

Type of Error	Data	Incorrect Category	OMB Category
Misclassify race	Country of origin = Iran	Asian	White
	Country of origin = Middle East	Asian	White
	Country of origin = European	Other	White
Check All that Apply	Mark Black category and in text specify AmerIndian	Black & Text	Black & Native American
Filipino	Category = Pacific Islander	Pacific Islander	Asian

Treva also provided reports on REDCap demographics, sample sizes and metrics of success. Table 4 shows there are data on nearly 1,300 DAP participants with 715 of them (about 55%) having one or more follow-up visits. The female:male ratio is about 3:2. Table 5 shows the race/ethnicity distribution of the aggregate sample. About 50% of the participants are Black or African American, over 36% are Hispanic or Latino, and the remaining 14% of the participants from underrepresented minority groups are Pacific Islanders, Native Americans or report multiple racial/ethnic backgrounds.

Table 4. Aggregate REDCap Demographics for the Diversity Action Plan (DAP) Programs

Aggregate Demographics

Program Sample Sizes		Miss	K12	K12/UG	UG	PB	Grad	PD	MF	Fac	Total
Target Sample Size (as of 3/4/13)											
Total Baseline		10	213	151	766	71	70	10	3	5	1,299
Follow-up Individuals		0	168	115	339	54	32	1	3	3	715
Follow-up Records		0	244	177	524	101	45	2	3	3	1,099
Lost for Follow-up	Deceased										0
	Lost Contact		24	2	10	3					39
	Refused		6	1	1						8
Female		3	127	92	447	46	37	5	2	3	762
Male		3	86	59	319	25	33	5	1	2	533
Missing Sex		4									4

Table 5. Aggregate REDCap Sample Sizes by Race and Ethnicity for the Diversity Action Plan (DAP) Programs

Aggregate URM Sample Sizes

	Missing	K-12	K-12/UG	UG	PB	Grad	PD	MF	Fac	Total
Undefined	5		10	20		5				40
NOT Underrepresented Minority Groups										
White	1		3	32	1					37
Asian			8	70	1	7				86
Mixed					1					1
Subtotal	1		11	102	3	7				124
Underrepresented Minority Groups										
Black	1	94	47	371	38	12	5	2	4	574
Am Ind		5	9	18	1	7	1			41
Pac Isl		3	6	8		4				21
Hispanic	2	98	55	208	17	29	3	1	1	414
>1 race	1	13	13	39	12	6	1			85
Subtotal	4	213	130	644	68	58	10	3	5	1135
Total	10	213	151	766	71	70	10	3	5	1299

The DACC highlighted several issues with data and data reporting:

(1) *Side Door Certifications No Longer Needed!* Now, you may enter your data from any computer without having it registered, as long as you know your login and password.

(2) *Missing Data:* Treva discussed the issue of missing data. If we are missing data then measuring program success becomes difficult. Missing data is an issue for general/baseline data (i.e. dates, race/ethnicity, gender, education, etc.) and follow-up data (i.e. publications, grants, current education/employment status, etc.).

(3) *Other questions to be considered by the programs and possible topic of discussion for October 2013 meeting:*
    i.   How often should each program follow-up with trainees/alumni?
    ii. Full-time vs. Part-time (defined during the training period)
    iii. Defining if trainees/alumni remain in the STEM field or research
    iv. How to define success that is appropriate for each education level?

(4) The DACC will consider allowing direct upload of data from existing databases into REDCap; Baylor College of Medicine and University of Washington are in pilot phases.

(5) *New rollouts coming out of the DACC soon include:* 1) missing data forms to PIs; 2) data correction forms to PIs (note: the DACC will NOT make any changes to the REDCap entries unless approved by PI/program coordinator); and 3) building new grants and publications databases from public sources. These rollouts will be a focus at the October 2013 training meeting.

(6) *Regular Notifications*: The DACC routinely sends notifications to programs reminding them of IRB expiration, grant expiration/renewal, and quarterly and annual data entry reminders for new trainees and alumni. Typically, a 2-month advance notice is provided, with a 1-month advance follow-up if needed.

(7) *Table of Common Data Problems / Data Queries:* In addition to regular notification, the DACC will query the programs for one or more of the following as needed.

Topic	Problem or Query
Awards	Define if your data reflect "submitted" or "received"
Target Sample	Define, separately for each career level, as total number ever trained
Duration	Define duration (# weeks) in terms of Summer or Academic year
Follow-up records	Goal is for at least 1 follow-up record per individual, even if refused follow-up, deceased, or lost contact. For refused, follow-up should only provide trainee id and note "refused"
ID/Name Key	Programs that have not supplied names and identifying information should provide this to the DACC as a "key". This information is needed for future follow-up, for creating publication and grants databases, and for matching IDs across programs. In order to be able to provide key to DACC, you may need to revisit your IRB if you fall into this category.

(8) *Table of Future Modifications to the REDCap Database:* Based on discussion and feedback in the data reports, the following modification will be made to the REDCap database.

Topic	Modification
Gender	Add transgender as a separate gender category
Date Finish	Distinguish among date (a) finished training program, (b) completed degree (c) completed association with program
Follow-up participation	Allow for multiple categories of participation such as (a) no longer associated with or supported by training program (b) not supported by training grant but still associated with training program (c) still supported on training grant
Awards	Distinguish between (a) total number SUBMITTED and (b) total number RECEIVED
Drop-down menus	New menus for several variables such as type of award, funding agency, etc
Publications	Define each as work (a) supported by this grant (b) not supported by this grant (c) unknown if supported by this grant
Termination	For T32, add "termination of appointment" form verbatim
Program type	For T32, should be either "combined" or "pre-doc", although trainees still tallied as either pre- or post-doc in statistics

(9) *List of Institutions/Programs and IRB status:* A list of each institution and their IRB status will be sent to each program by separate e-mail. This list indicates whether the program has obtained an exempt or expedited status, whether

informed consents are required, and if a 'Key' is required for full identification information on participants.

IV.   MEETING SUMMARY

a. General Comments

- IRB: All attendees agreed that IRB status and documentation could be shared among programs. Context: Having access to example IRB applications that requested "waiver" versus annual review.
- Multiples: Trainees who participate across different programs, and across different institutions.
  - DACC performs regular checks to identify these individuals. This is a problem if DACC does not have name (i.e. only ID)
  - Who performs follow-up?
    - If multiple programs are within same institution, then there is no problem in determining who will perform follow-up.
    - If multiple programs are across different institutions, then discuss among DACC and both institutions to determine who will follow-up (one or both).
    - REDCap: data access groups restrict viewing. Will need to open views across these two institutions for a given trainee on a case-by-case basis.
- Other:
  - Individual Development Plan (IDP) or Career Development Plan (CDP). May be required soon by many programs.
  - URM and work permits. Even if highly qualified, undocumented individuals cannot be supported on grants. Dream Act currently under discussion may change this?
  - Crash in R01 funding is becoming a problem. For example, student may have funding but the lab is no longer funded. NIH is pushing for F31s. Merna suggests taking F31 to use for extending existing funding.

b. Challenges Faced
- **Follow-up / tracking**
  - Follow-up data collection
  - Tracking students
  - How to keep less compliant alumni engaged for follow-up
  - Increased use of social media and web-based resources
  - Getting written consents from trainees no longer at UWM
  - Follow-up: Successful alumni are more likely to respond than those who change career focus or who are less successful (try to maintain contact

more frequent than once a year, stay in contact with dissertation advisors, investigation of websites and publication searches)

- o **Recruiting and Competition for qualified trainees**
  - Admissions committees misunderstanding CA Prop 209
  - Lack of URM students in computation fields such as bioinformatics
  - Recruiting more URM
  - Number of URM prepared for genomics study at doctoral level is relatively small, but growing; competition among top programs is intense
  - Ensuring good balance between biology and quantitative skills, having mentors from both disciplines helps
  - Finding trainees with balance between biology and quantitative skills
  - Recruiting problems, does not have "brand" name of large universities
  - Attraction to industry jobs in region
  - Fierce competition for small pool of talented URM PhD candidates (cast wider net, education admissions committee, coordinate with other DAP programs)
  - Identifying strong candidates
  - Faculty poaching from private institutions
  - Identifying students who will be mature enough two summers later
  - Database projects may not appeal to those wanting laboratory research
  - Adjustment to isolation / rural environment

- o **Funding**
  - Have more students than can fund
  - Funding
  - Intense R01 funding competition
  - $ pressure on trainees
  - Job market is difficult for graduates, especially searching for faculty positions

- o **Internal Resources**
  - Finding new laboratories
  - Developing bioinformatics boot camp
  - Psychological assessments needed
  - Connecting students across range of departments, programs, campuses
  - Mentoring

- o **Data Entry**
  - Translating informal report of outcomes (phone calls, texts, etc) and standardize to REDCap

V. MINI GRANTS WRITING & GRANTS MANAGEMENT SEMINAR

   a. Grants Writing

Bettie Graham gave a presentation entitled, "Helpful Tips for Trainees and New Investigators Seeking Support from the National Institutes of Health".   Bettie walked through the grant application life cycle processes of application submission, review, funding decisions and award. When considering submission of an F or K application, the key ingredients to getting started are to: have a good idea; check that your idea is sufficiently unique (NIH RePORT, literature search); know the research interests of the NIH ICs; and very importantly, *talk to an NIH Program Officer early in the concept development stage of the application.* Other take home messages are to allow plenty of time to write/review/rewrite; and to not rely solely on your mentor, department chair or friend to interpret NIH program needs. A list of resources was provided.

   b. Grants Management

Susan Toy gave an overview of grants management of training and career development awards.   Susan reviewed award budgets, project periods, indirect thresholds, and allowable/unallowable costs.   Other aspects of pre-award review, award management and closeout were addressed.

ACKNOWLEDGEMENT: A special thanks to Carolyn Taylor for her help with organizing the workshop.

#1 NOTE: On Tracking Responsibilities

Training coordinators are required to track participants for ten years. This has been and will always be their responsibility as long as they receive funding from NHGRI. This is NOT a DACC responsibility until after the ten-year period. The big issue comes when grants are terminated; in this case the expectation is that the DACC will have access to the original records on past participants and will be able to track them.

Whereas our focus is to get a profile on current trainees, the expectation is that training coordinators will provide records of past participants. It may not be possible to have complete information on past participants, but training coordinators have always had a requirement to track trainees, and thus, should have some records.

#2 NOTE: The date/place for the annual Fall DAP Meeting has been confirmed: October 14-15 at the Fluno Center in Madison, WI; hosted by University of Wisconsin-Madison.

FINAL AGENDA

2013 TRAINING COORDINATORS' WORKSHOP

Fifth Floor Conference Room
5625 Fishers Lane
Rockville, MD 20852

22 April 2013

Purpose of Meeting: (1) to review progress toward meeting the goal of increasing the number of URMs in genomic sciences; (2) to review the quality of the data submitted to REDCap; and (3) to provide a mini-grant writing and grants management workshop.

8:30 a.m.   **Welcome and Introductions**
        Bettie J. Graham

8:50   **Progress Reports—Institutional Training Grants**
        (10 minutes presentation; 10 minutes discussion)
        Chair: Tina Gatlin

        University of California, Berkeley
        University of California, Los Angeles
        Harvard
        Princeton University
        Stanford University

10:50   **Break**

11:20   **Progress Reports—Institutional Training Grants--Continued**
        University of Michigan
        University of Washington
        University of Wisconsin
        Yale University
        University of Pennsylvania
        Massachusetts Institute of Technology
        Washington University in St. Louis

1:00   **Lunch (on your own, cafeteria next door)**

2:00   **REDCap Data Report and Discussion**
        (Treva Rice and DACC)

2:45   **Break**

3:15   **Mini Grants Writing and Grants Management Seminar**
        (Susan Toy and Bettie J. Graham)

5:15   **Participant Generated Discussion**

6:00   **Adjourn**

23 April 2013

8:30   **Progress Reports--DAPs Associated with Large-Scale Sequencing, Database and CEGS Grants**

Chair: Heather Junkins

Baylor College of Medicine
The Broad Institute
Washington University, St. Louis
University of California, Santa Cruz
Harvard University/University of New Mexico
The Jackson Laboratory

10:30   **Break**

11:00   Dana Farber Cancer Institute
Harvard University
Johns Hopkins University
University of Southern California
University of Washington

12:40   **Lunch (on your own, cafeteria next door)**

1:40   **REDCap Report and Discussion**
(Treva Rice and DACC) and Dawayne Whittington

2:40   **General Discussion**

3:15   **Adjourn**

NOTE: Fall meeting will be in Madison, WI on 14-15 October.

NHGRI Research Training Coordinators' Meeting
5625 Fishers Lane
Rockville, MD
April 22-23, 2013

RESEARCH TRAINING ADVISORY COMMITTEE

Merna Villarejo
University of California, Davis
2530 Whittier Drive
Davis, CA 95618
(530) 756-2342
mrvillarejo@ucdavis.edu

Kim J. Nickerson
University of Maryland, College Park
2141 Tydings Hall,
College Park, MD 20742
(301) 405-7599
knick@umd.edu

CENTERS OF EXCELLENCE IN GENOMIC SCIENCE (CEGS)

Karen Burns White
Dana-Farber Cancer Institute
Harvard Cancer Center
450 Brookline Avenue, BP344
Boston, MA, 02215
(617) 632-3244
Karen_burnswhite@dfci.harvard.edu

Alexander Hernandez-Siegel
Harvard Medical School
Department of Genetics
New Research Building, Room 238H, Church Lab
77 Avenue Louis Pasteur
Boston, MA 02115
(617) 432-5742
ahs@genetics.med.harvard.edu

David Hill
Harvard Medical School
44 Binney Street
Boston, MA 02115
(617) 632-3802
David_hill@dfci.harvard.edu

Steve Finkel
University of Southern California
3551 Trousdale Pkwy.
ADM 304
Los Angeles, CA 90089-4012
(213) 821-1498
sfinkel@usc.edu

Allison Kang
University of Washington
Box 352180
Seattle, WA 98195-2120
(206) 221-3056
allikang@u.washington.edu

Lisa Peterson
University of Washington
Box 352180
Seattle, WA 98195-2120
(206) 685-2593
Lisapete@u.washington.edu

Vicky Schneider
Johns Hopkins University
5801 Smith Avenue
McAuley Hall, STE 400
Baltimore, MD 21209
(410) 735-6219
vschneider@jhu.edu

Latishya Steele
Dana-Farber Cancer Institute
Research Administration
Harvard Medical School
450 Brookline Ave , BP322A
Boston, MA 02115
(617) 632-3028
Latishya_steel@dfci.harvard.edu

LARGE SCALE SEQUENCING

Francie Latour
Broad Institute
301 Binney Street
Cambridge, MA 02142
(617) 714-7000
diversity@broadinstitute.org

Debra Murray
Baylor College of Medicine
One Baylor Plaza N1519
Houston, TX 77030
(713) 798-8083
ddm@bcm.edu

Cherilynn Shadding
Washington University in St. Louis-SOM
4444 Forest Park Blvd
Campus Box 8501
St. Louis, MO 63108
(314) 286- 1897
cshadding@genome.wustl.edu

Eboney Smith
Broad Institute
7 Cambridge Center
Cambridge, MA 02142
(617) 324-1237
esmith@broad.mit.edu

INSTITUTIONAL TRAINING GRANTS

Anita Blanco
Stanford University
300 Pasteur Drive
Alway Building, Room M350
Stanford, CA 94305-5120
Phone: (650) 723-6274
E-mail: anita.blanco@stanford.edu

Susanne Churchill
National Center for Biomedical Computing
77 Avenue Louis Pasteur
Boston, MA 02115
(617) 525-4465
schurchill@partners.org

Jeanne Darling
MIT
32 Vassar Street
Cambridge, MA 02139
(617) 253-4294
darling@mit.edu

Katherine Flannery
Harvard Medical School
10 Shattuck Street
Boston, MA 02115
617.432.7294
Katherine_Flannery@hms.harvard.edu

Jennifer Frederick
Yale University
PO Box 208103
219 Prospect Street, KBT 950
New Haven, CT 06520
203.432.3515
jennifer.frederick@yale.edu

Alison Gammie
Princeton University
334 Lewis Thomas Laboratory
Princeton, NJ 08544
(609) 258-6380
agammie@princeton.edu

Jo Handelsman
Yale University
219 Prospect Street
New Haven, CT 06520
(203) 432-9119
jo.handelsman@yale.edu

Dawn Keene
University of Michigan
1415 Washington Heights, M4242
Ann Arbor, MI 48109
(734) 647-3944
dhke@umich.edu

Louise Pape
University of Wisconsin-Madison
3445 Biotechnology Center
425 Henry Mall
Madison WI 53706
(608) 265-7935
lpape@wisc.edu

Jeanette Papp
University of California, Los Angeles
695 Charles E. Young Drive South
Los Angeles, CA 90095-7088
(310) 825-6204
jcpapp@mednet.ucla.edu

Jasper Rine
University of California Berkeley
374A Stanley Hall
Berkeley, CA 94720
510-642-7047
jrine@berkeley.edu

David C. Schwartz
University of Wisconsin-Madison
425 Henry Mall
Madison, WI 53706
(680) 265-0546
dcschwartz@wisc.edu

Lyle Ungar
University of Pennsylvania
220 South 33rd Street
107 Towne Building
Philadelphia, PA 19104-6391
(215) 898-7246
ungar@cis.upenn.edu

DATABASES

Joanne Berghout
Mouse Genome Informatics
600 Main Street
Bar Harbor, ME 04609
(207) 288-6426
joanne.berghout@jax.org

William Gelbart
Harvard University
16 Divinity Avenue, Room 4059
Cambridge, MA 02138
(617) 495-2906
Gelbart@morgan.harvard.edu

Zia Isola
University of California, Santa Cruz
1156 High Street, CBSE-ITI,
Santa Cruz, CA 95064
(813) 459-1702
isola@soe.ucsc.edu

James Kent
University of California Santa Cruz
156 high Street
Santa Cruz, CA 95064
(831-459-5127
kent@soe.ucsc.edu

Paul Szauter
University of New Mexico
167 Castetter Hall, MSC03 2020
Albuquerque, NM 87131 0001
(207) 812-8932
PSzauter@unm.edu

DATA ANALYSIS AND COORDINATING CENTER (DACC)

Jeanne Cashman
Washington University School of Medicine
660 South Euclid Avenue, Box 8067
St. Louis, MO 63110
(314) 362-3615
jeanne@wubios.wustl.edu

Donna Jeffe
Washington University School of Medicine
4444 Forest Park Blvd Box 8504
St. Louis, MO 63108
Djeffe@dom.wustl.edu

Karen Clark Laseter
Washington University School of Medicine
660 South Euclid Avenue Box 8067
St. Louis, MO 63110
(314) 362-2349
karen@wubios.wustl.edu

Treva Rice
Washington University School of Medicine
660 South Euclid Avenue Box 8067
St. Louis, MO 63110
(314) 362-3662
treva@wubios.wustl.edu

NATIONAL HUMAN GENOME RESEARCH INSTITUTE STAFF

Carla L. Easter
National Human Genome Research Institute
National Institutes of Health
Bldg 31, Room B1B55, MSC 2070
31 Center Drive
Bethesda MD 20892-2070
(301) 594-1364
esterc@mail.nih.gov

Christine Gatlin
National Human Genome Research Institute
National Institutes of Health
5635 Fishers Lane, Suite 4076/MSC 9305

Rockville, MD 20852-9305
(301) 496-7531
christine.gatlin@nih.gov

Bettie J. Graham
National human Genome Research Institute
National Institutes of Health
5635 Fishers Lane, Suite 4076
Bethesda, MD 20892-9305
(301) 496-7531
Bettie_graham@nih.gov

Eric D. Green
Director
National Human Genome Research Institute
National Institutes of Health
31 Center Drive
Building 31, Room 4B09
Bethesda, MD 20892
(301) 496-0844
egreen@nhgri.nih.gov

Mark Guyer
National Human Genome Research Institute
National Institutes of Health
5635 Fishers Lane, Suite 4076
Bethesda, MD 20892-9305
(301) 496-7531
Mark_guyer@nih.gov

Faith Harrow
NHGRI Training Program Coordinator
NHGRI Intramural Training Office
Building 12A, Room 1013, 123 South Drive
Bethesda, MD 20892-5613
(301) 451-3645
harrowf@mail.nih.gov

Heather Junkins
National Human Genome Research Institute
National Institutes of Health
5635 Fishers Lane, suite 4076/MSC 9305
Bethesda, MD 20892-9305
(301) 496-7531
heather.junkins@nih.gov

Lisa Oken
National Human Genome Research Institute
National Institutes of Health
5635 Fishers Lane, suite 4076/MSC 9305
Bethesda, MD 20892-9305
(301) 496-7531
loken@mail.nih.gov

William Pavan
National Human Genome Research Institute
National Institutes of Health
49 Convent Drive
Bethesda, MD 20892
(301) 496-7584
William.Pavan@nih.gov

Jeffery A. Schloss
National Human Genome Research Institute
National Institutes of Health
5635 Fishers Lane, Suite 4076
Bethesda, MD 20892-9305
(301) 496-7531
Jeff_schloss@nih.gov

Carolyn Taylor (Technical Assistance)
National Human Genome Research Institute
National Institutes of Health
5635 Fishers Lane, Suite 4076/MSC
Rockville, MD 20852-9305
(301) 496-7531
taylorca@mail.nih.gov

Susan Toy
National Human Genome Research Institute
National Institutes of Health
5635 Fishers Lane, suite 4076/MSC 9305
Bethesda, MD 20892-9305
(301) 496-7531
toys@mail.nih.gov