PI: **Boehm, Frederick J**	Title: Sample size planning methods in post-GWAS breast cancer eQTL studies	
Received: 12/10/2012	FOA: PA11-113	Council: 05/2013
Competition ID: ADOBE-FORMS-B1	FOA Title: RUTH L. KIRSCHSTEIN NATIONAL RESEARCH SERVICE AWARDS (NRSA) FOR INDIVIDUAL POSTDOCTORAL FELLOWS (PARENT F32)	
1 F32 CA180655-01	Dual:	Accession Number: 3549562
IPF: 578503	Organization: UNIVERSITY OF WISCONSIN-MADISON	
Former Number:	Department: Biostatistics/Med Info	
IRG/SRG: ZRG1 F09A-L (20)L	AIDS: N	Expedited: N
Subtotal Direct Costs (excludes consortium F&A)	Animals: N Humans: N Clinical Trial: N Current HS Code: 10 HESC: N	New Investigator: N Early Stage Investigator: N
Senior/Key Personnel:	*Organization:*	*Role Category:*
Frederick Boehm	The Board of Regents of the University of Wisconsin System	PD/PI
MICHAEL NEWTON	The Board of Regents of the University of Wisconsin System	Other (Specify)-Sponsor
MICHAEL GOULD	The Board of Regents of the University of Wisconsin System	Other (Specify)-Co-Sponsor

Reference Letters

Bruce Weir	University of Washington	12/10/2012
Chunming Zhang	University of Wisconsin-Madison	12/10/2012
Christina Newton	UW-Madison	12/10/2012
David Crosslin	University of Washington	12/10/2012

OMB Number: 4040-0001
Expiration Date: 06/30/2011

APPLICATION FOR FEDERAL ASSISTANCE
SF 424 (R&R)

3. DATE RECEIVED BY STATE	State Application Identifier

1. * TYPE OF SUBMISSION
☐ Pre-application ☒ Application ☐ Changed/Corrected Application

2. DATE SUBMITTED: 12/10/2012

Applicant Identifier:

4. a. Federal Identifier:

b. Agency Routing Identifier:

5. APPLICANT INFORMATION

*** Organizational DUNS:** 161202122

* Legal Name: The Board of Regents of the University of Wisconsin System

Department: | Division:

* Street1: Suite 6401

Street2: 21 N Park St

* City: Madison | County / Parish: Dane

* State: WI: Wisconsin | Province:

* Country: USA: UNITED STATES | * ZIP / Postal Code: 53715-1218

Person to be contacted on matters involving this application

Prefix: | * First Name: DEBORAH | Middle Name: M

* Last Name: MELTZER | Suffix:

* Phone Number: 6082634940 | Fax Number: 6082626565

Email: DMELTZER@WISC.EDU

6. * EMPLOYER IDENTIFICATION *(EIN) or (TIN):* 396006492

7. * TYPE OF APPLICANT: H: Public/State Controlled Institution of Higher Education

Other (Specify):

Small Business Organization Type ☐ Women Owned ☐ Socially and Economically Disadvantaged

8. * TYPE OF APPLICATION:
☒ New ☐ Resubmission
☐ Renewal ☐ Continuation ☐ Revision

If Revision, mark appropriate box(es).
☐ A. Increase Award ☐ B. Decrease Award ☐ C. Increase Duration ☐ D. Decrease Duration
☐ E. Other (specify):

* Is this application being submitted to other agencies? Yes ☐ No ☒ What other Agencies?

9. * NAME OF FEDERAL AGENCY: National Institutes of Health

10. CATALOG OF FEDERAL DOMESTIC ASSISTANCE NUMBER:

TITLE:

11. * DESCRIPTIVE TITLE OF APPLICANT'S PROJECT:
Sample size planning methods in post-GWAS breast cancer eQTL studies

12. PROPOSED PROJECT:
* Start Date: 07/01/2013 | * Ending Date: 01/31/2015

*** 13. CONGRESSIONAL DISTRICT OF APPLICANT**: WI-002

14. PROJECT DIRECTOR/PRINCIPAL INVESTIGATOR CONTACT INFORMATION

Prefix: | * First Name: Frederick | Middle Name: J

* Last Name: Boehm | Suffix:

Position/Title: POSTDOCTORAL TRAINEE

* Organization Name: The Board of Regents of the University of Wisconsin System

Department: Biostatistics/Med Info | Division: SCHOOL OF MEDICINE AND PUBLIC

* Street1: 5730 Medical Sciences Center

Street2: 1300 University Ave

* City: Madison | County / Parish: Dane

* State: WI: Wisconsin | Province:

* Country: USA: UNITED STATES | * ZIP / Postal Code: 53706-1510

* Phone Number: 608-890-1248 | Fax Number:

* Email: fjboehm@wisc.edu

15. ESTIMATED PROJECT FUNDING

a. Total Federal Funds Requested	155,744.00
b. Total Non-Federal Funds	0.00
c. Total Federal & Non-Federal Funds	155,744.00
d. Estimated Program Income	0.00

16. * IS APPLICATION SUBJECT TO REVIEW BY STATE EXECUTIVE ORDER 12372 PROCESS?

a. YES ☐ THIS PREAPPLICATION/APPLICATION WAS MADE AVAILABLE TO THE STATE EXECUTIVE ORDER 12372 PROCESS FOR REVIEW ON:

DATE:

b. NO ☒ PROGRAM IS NOT COVERED BY E.O. 12372; OR

☐ PROGRAM HAS NOT BEEN SELECTED BY STATE FOR REVIEW

17. By signing this application, I certify (1) to the statements contained in the list of certifications* and (2) that the statements herein are true, complete and accurate to the best of my knowledge. I also provide the required assurances * and agree to comply with any resulting terms if I accept an award. I am aware that any false, fictitious. or fraudulent statements or claims may subject me to criminal, civil, or administrative penalities. (U.S. Code, Title 18, Section 1001)

☒ *** I agree**

**** The list of certifications and assurances, or an Internet site where you may obtain this list, is contained in the announcement or agency specific instructions.***

18. SFLLL or other Explanatory Documentation

Add Attachment | Delete Attachment | View Attachment

19. Authorized Representative

Prefix: * First Name: NICHOLAS Middle Name: N

* Last Name: NOVAK Suffix:

* Position/Title: MANAGING OFFICER

* Organization: The Board of Regents of the University of Wisconsin System

Department: RESEARCH AND SPONSORED PROGRAM Division:

* Street1: 21 N Park Street

Street2: Suite 6401

* City: MADISON County / Parish: Dane

* State: WI: Wisconsin Province:

* Country: USA: UNITED STATES * ZIP / Postal Code: 53715-1218

* Phone Number: 608-262-3822 Fax Number: 608-262-5111

* Email: PREAWARD@RSP.WISC.EDU

*** Signature of Authorized Representative**: NICHOLAS N NOVAK

*** Date Signed**: 12/10/2012

20. Pre-application

Add Attachment | Delete Attachment | View Attachment

424 R&R and PHS-398 Specific Table Of Contents

OMB Number: 4040-0010
Expiration Date: 08/31/2011

Project/Performance Site Location(s)

Project/Performance Site Primary Location ☐ I am submitting an application as an individual, and not on behalf of a company, state, local or tribal government, academia, or other type of organization.

Organization Name: The Board of Regents of the University of Wisconsin System

DUNS Number: 1612021220000

* Street1: Suite 6401

Street2: 21 N Park St

* City: Madison County: Dane

* State: WI: Wisconsin

Province:

* Country: USA: UNITED STATES

* ZIP / Postal Code: 53715-1218 * Project/ Performance Site Congressional District: WI-002

Project/Performance Site Location 1 ☐ I am submitting an application as an individual, and not on behalf of a company, state, local or tribal government, academia, or other type of organization.

Organization Name:

DUNS Number:

* Street1:

Street2:

* City: County:

* State:

Province:

* Country: USA: UNITED STATES

* ZIP / Postal Code: * Project/ Performance Site Congressional District:

Additional Location(s) Add Attachment Delete Attachment View Attachment

RESEARCH & RELATED Other Project Information

1. * Are Human Subjects Involved? ☐ Yes ☒ No

1.a If YES to Human Subjects

Is the Project Exempt from Federal regulations? ☐ Yes ☐ No

If yes, check appropriate exemption number. ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6

If no, is the IRB review Pending? ☐ Yes ☐ No

IRB Approval Date:

Human Subject Assurance Number:

2. * Are Vertebrate Animals Used? ☐ Yes ☒ No

2.a. If YES to Vertebrate Animals

Is the IACUC review Pending? ☐ Yes ☐ No

IACUC Approval Date:

Animal Welfare Assurance Number

3. * Is proprietary/privileged information included in the application? ☐ Yes ☒ No

4.a. * Does this project have an actual or potential impact on the environment? ☐ Yes ☒ No

4.b. If yes, please explain:

4.c. If this project has an actual or potential impact on the environment, has an exemption been authorized or an environmental assessment (EA) or environmental impact statement (EIS) been performed? ☐ Yes ☐ No

4.d. If yes, please explain:

5. * Is the research performance site designated, or eligible to be designated, as a historic place? ☐ Yes ☒ No

5.a. If yes, please explain:

6. * Does this project involve activities outside of the United States or partnerships with international collaborators? ☐ Yes ☒ No

6.a. If yes, identify countries:

6.b. Optional Explanation:

7. * Project Summary/Abstract projectSummary_v41013457497.pdf Add Attachment | Delete Attachment | View Attachment

8. * Project Narrative projectNarrative_v31013457496.pdf Add Attachment | Delete Attachment | View Attachment

9. Bibliography & References Cited bibRefCited_v21013457495.pdf Add Attachment | Delete Attachment | View Attachment

10. Facilities & Other Resources facilities2012_F321013448839.pdf Add Attachment | Delete Attachment | View Attachment

11. Equipment Add Attachment | Delete Attachment | View Attachment

12. Other Attachments Add Attachments | Delete Attachments | View Attachments ☐

Project summary

Frederick Boehm, M.D. seeks to translate genome-wide association study (GWAS) results into clinical applications and to become an independent biomedical investigator. To achieve these goals, he will undertake a rigorous postdoctoral training program in biostatistical genomics under the mentorship of Michael Newton and Michael Gould. The research component of Boehm's program requires him to: 1) develop methods for expression quantitative trait locus (eQTL) study sample size planning with false discovery rate control in the setting of post-GWAS SNP characterization, and 2) develop sample size planning methods for pathway-based inference in eQTL studies. Boehm's approach uses hierarchical mixture modeling of variations in gene expression levels. To assess the statistical models, Boehm will apply his methods both in simulation studies and in analysis of pilot genomics data from 62 breast tissue samples.

The proposed research will enable post-GWAS SNP characterization studies. Such studies are a critical intermediate step between GWAS results and their clinical application. These post-GWAS studies have the potential for major public health impacts by enabling personalized genomic medicine.

Boehm's training plan includes courses in statistical theory and methods. Course topics include Bayesian methods, high-dimensional inference, survival analysis, and statistical consulting. These courses also will fulfill a requirement for the statistics Ph.D. program in which Boehm is a student.

Boehm will supplement formal coursework with independent study of key texts and journal articles in multivariate and high-dimensional statistics. Professor Newton will meet regularly with Boehm to discuss questions that arise from his readings.

To facilitate professional development throughout the fellowship period, Boehm will enroll in courses that address skills such as grant proposal writing, manuscript writing, research team management, and research mentoring. To refine mentoring skills, he will train a summer undergraduate researcher during his second year. Boehm also will undertake an ambitious program for training in responsible conduct of research that includes a capstone research ethics presentation to the university community.

Project narrative

Our project will develop biostatistical tools to advance public health by enabling 1) discovery of novel biological targets for drugs to treat breast cancer and 2) development of markers for early detection of breast cancer. In this manner, our research is a translational bridge between recent advances in genetics technology and improvements in breast cancer diagnosis and treatment.

BIBLIOGRAPHY AND REFERENCES CITED

1. Hindorff LA, MacArthur J, Morales J, Junkins HA, Hall PN, Klemm AK, Manolio TA. *A Catalog of Published Genome-Wide Association Studies.* at <http://www.genome.gov/gwastudies/>
2. Freedman, M. L., Monteiro, A. N. A., Gayther, S. A., Coetzee, G. A., Risch, A., Plass, C., Casey, G., De Biasi, M., Carlson, C., Duggan, D. & others. Principles for the post-GWAS functional characterization of cancer risk loci. *Nature genetics* **43,** 513–518 (2011).
3. Emilsson, V., Thorleifsson, G., Zhang, B., Leonardson, A. S., Zink, F., Zhu, J., Carlson, S., Helgason, A., Walters, G. B., Gunnarsdottir, S. & others. Genetics of gene expression and its effect on disease. *Nature* **452,** 423–428 (2008).
4. Nicolae, D. L., Gamazon, E., Zhang, W., Duan, S., Dolan, M. E. & Cox, N. J. Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS genetics* **6,** e1000888 (2010).
5. Gerlinger, M., Rowan, A. J., Horswell, S., Larkin, J., Endesfelder, D., Gronroos, E., Martinez, P., Matthews, N., Stewart, A., Tarpey, P. & others. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *New England Journal of Medicine* **366,** 883–892 (2012).
6. Cowper-Sal, R., Zhang, X., Wright, J. B., Bailey, S. D., Cole, M. D., Eeckhoute, J., Moore, J. H., Lupien, M. & others. Breast cancer risk-associated SNPs modulate the affinity of chromatin for FOXA1 and alter gene expression. *Nature Genetics* (2012).
7. McCarthy, M. I., Abecasis, G. R., Cardon, L. R., Goldstein, D. B., Little, J., Ioannidis, J. P. A. & Hirschhorn, J. N. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nature Reviews Genetics* **9,** 356–369 (2008).
8. Knuth, D. E. Literate programming. *The Computer Journal* **27,** 97–111 (1984).
9. Leisch, F. Sweave, Part I: Mixing R and \LaTeX. *R News* **2,** 28–31 (2002).
10. Gentleman, R. C., Carey, V. J., Bates, D. M. & others. Bioconductor: Open software development for computational biology and bioinformatics. *Genome Biology* **5,** R80 (2004).
11. *Run My Code*. at <http://runmycode.org>
12. Laurie, C. C., Doheny, K. F., Mirel, D. B., Pugh, E. W., Bierut, L. J., Bhangale, T., Boehm, F., Caporaso, N. E., Cornelis, M. C., Edenberg, H. J. & others. Quality control and quality assurance in genotypic data for genome-wide association studies. *Genetic epidemiology* **34,** 591–602 (2010).
13. Team, R. D. C. *R: A Language and Environment for Statistical Computing*. (2008). at <http://www.R-project.org>
14. Nica, A. C., Montgomery, S. B., Dimas, A. S., Stranger, B. E., Beazley, C., Barroso, I. & Dermitzakis, E. T. Candidate causal regulatory effects by integration of expression QTLs with complex trait genetic associations. *PLoS genetics* **6,** e1000895 (2010).
15. Sillé, F. C. M., Thomas, R., Smith, M. T., Conde, L. & Skibola, C. F. Post-GWAS Functional Characterization of Susceptibility Variants for Chronic Lymphocytic Leukemia. *PloS one* **7,** e29632 (2012).
16. Wallace, C., Rotival, M., Cooper, J. D., Rice, C. M., Yang, J. H. M., McNeill, M., Smyth, D. J., Niblett, D., Cambien, F., Tiret, L. & others. Statistical colocalization of monocyte gene expression and genetic risk variants for type 1 diabetes. *Human molecular genetics* **21,** 2815–2824 (2012).
17. Newton, M., Wang, P. & Kendziorski, C. Hierarchical mixture models for expression profiles. *Bayesian inference for gene expression and proteomics* 40–52 (2006).
18. Ji, H. & Liu, X. S. Analyzing'omics data using hierarchical models. *Nature biotechnology* **28,** 337 (2010).
19. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)* 289–300 (1995).
20. Storey, J. D. A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64,** 479–498 (2002).
21. Storey, J. D. The positive false discovery rate: A Bayesian interpretation and the q-value. *Annals of Statistics* 2013–2035 (2003).
22. Jørstad, T. S., Midelfart, H. & Bones, A. M. A mixture model approach to sample size estimation in

two-sample comparative microarray experiments. *BMC bioinformatics* **9,** 117 (2008).
23. Hu, J., Zou, F. & Wright, F. A. Practical FDR-based sample size calculations in microarray experiments. *Bioinformatics* **21,** 3264–3272 (2005).
24. Borman, S. *The Expectation Maximization Algorithm-A short tutorial*. (2009).
25. *Center for High Throughput Computing*. at <http://chtc.cs.wisc.edu/>
26. Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H. & Kanehisa, M. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic acids research* **27,** 29–34 (1999).
27. Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T. & others. Gene Ontology: tool for the unification of biology. *Nature genetics* **25,** 25 (2000).
28. Joshi-Tope, G., Gillespie, M., Vastrik, I., D'Eustachio, P., Schmidt, E., De Bono, B., Jassal, B., Gopinath, G., Wu, G., Matthews, L. & others. Reactome: a knowledgebase of biological pathways. *Nucleic acids research* **33,** D428–D432 (2005).
29. Newton, M. A., Quintana, F. A., Den Boon, J. A., Sengupta, S. & Ahlquist, P. Random-set methods identify distinct aspects of the enrichment signal in gene-set analysis. *The Annals of Applied Statistics* 85–106 (2007).
30. Newton, M. A., He, Q. & Kendziorski, C. A model-based analysis to infer the functional content of a gene list. (2011).
31. Newton, M. *allez: Random-set calibration of gene-set statistics*. (2009).

FACILITIES AND OTHER RESOURCES

THE UNIVERSITY OF WISCONSIN

The University of Wisconsin is one of the nation's leading and largest research institutions, well known, for example, for its early discovery of warfarin, the sequencing of the *E. Coli* genome, and the recent development of stem cell research. The University of Wisconsin at Madison (UW-Madison) is a large campus of 40,000 students with major colleges and schools of Letters and Science, Agricultural and Life Sciences, Veterinary Medicine, Medicine, Nursing, Law, Education, Pharmacy, Business, and Engineering with a budget of $1.6 billion. This large single campus provides an exceedingly rich research environment

THE UNIVERSITY OF WISCONSIN MEDICAL SCHOOL

The University of Wisconsin School of Medicine and Public Health, founded in 1907, has 1300 faculty appointed in 11 basic science departments and 14 clinical departments. The graduate program features over 650 students and there are 615 medical students. Total extramural research support is approximately $325 million dollars, which is nearly 30% of the total grants to the University of Wisconsin in 2010. Recognized as an international, national and statewide leader in educating physicians, investigating the causes of disease, exploring innovative solutions to medical problems and translating research into compassionate patient care, the UW School of Medicine and Public Health seeks to attract the very best students, educators and researchers in pursuit of our mission. Through programs that are interdisciplinary and translational, the UW School of Medicine and Public Health is at the vanguard of today's medicine and tomorrow's cures. The UW School of Medicine and Public Health offers students, educators and researchers access to all of the benefits of a preeminent public research university. In addition, the UW School of Medicine and Public Health has strong partnerships with the University of Wisconsin Hospital and Clinics, whose programs are consistently ranked among the best in the nation; and the University of Wisconsin Medical Foundation, one of the 10 largest physician practice groups in the country. These relationships afford students, teachers and researchers opportunities available only at top ranked academic medical centers.

UW DEPARTMENT OF BIOSTATISTICS AND MEDICAL INFORMATICS

Within the UW School of Medicine and Public Health, the Department of Biostatistics and Medical Informatics provides a vital environment for research, collaboration, and training in biostatistics and informatics. There are now 34 PhD-level statisticians and computer scientists with appointments (full or partial) in the Department, 7 having joint appointments in the Statistics Department and 14 having joint appointments in other departments such as Computer Science and Population Health Sciences. These strong faculty-level connections support a high level of interdisciplinary research and enable the coordination of graduate student training.

What is now the Department of Biostatistics and Medical Informatics originated in 1984 as the Biostatistics Center and became formally approved by the University in 1986 as the focus for statistical research and collaboration between the Statistics Department and the Medical School. Its development really began in 1982 with the recruitment of Dr. David DeMets from the NIH to develop a Cancer Biostatistics Program for the UW Clinical Cancer Center. The Biostatistics Center gained departmental status in 1992. In 1996 the name was changed to Biostatistics and Medical Informatics to reflect the growth and development in the rapidly evolving field of bioinformatics and information technology with research activities focused on bioinformatics, computational molecular biology, clinical informatics and medical imaging.

The Department of Biostatistics and Medical Informatics is organized into 3 major research and training programs: Biostatistics, Clinical Trials and Medical Informatics. The Department hosts an NIGMS predoctoral training grant in Biostatistics, and participates in a National Library of Medicine Training grant for Computation and Informatics in Biology and Medicine. The Department of Biostatistics and Medical Informatics currently employs over 80 individuals, including MS-level biostatisticians, programmers, and data managers.

Faculty members are principal investigators on a number of investigator-initiated research projects at NIH and NSF in areas such as: statistical methods for cancer research, statistical techniques for genomics; survival analysis, sequential methods for clinical trials, statistical methods for molecular cancer data, analysis of

functional data, demographic data analysis, adaptive information monitoring and extraction; machine learning with rich data sources; computational methods for identifying co-regulated genes from expression and text data; view synthesis for dynamic scenes with and without reconstruction; relational learning systems; relational pattern learning, and quality of life data analysis.

In addition to methodological research, the faculty in the Department of Biostatistics and Medical Informatics collaborate with clinical and laboratory investigators on campus, nationally and internationally. These collaborations often motivate biostatistics and bioinformatics research on new methods to answer scientific questions, further enhancing collaboration and scientific interaction.

Most faculty members in the Department of Biostatistics and Medical Informatics are located at the Clinical Science Center (CSC) in 2,000 square feet of space. The CSC is part of the Health Sciences complex on the west end of the Madison campus. The Department also has 5,054 square feet in the WARF Building next to the CSC for the Clinical Trials Program. In the Biological Basic Science complex, at the center of campus and about one block from the Statistics Department, the Department has another 5321 square feet in the Medical Sciences Center building for the Medical Informatics Program.

COMPUTING RESOURCES IN THE DEPARTMENT OF BIOSTATISTICS AND MEDICAL INFORMATICS

Over the past twenty years, the Department of Biostatistics and Medical Informatics has developed a centralized state-of-the-art computing facility for the support of statistical and medical informatics research, and for the management and analysis of clinical, genomic and other biological data.

The computational resources in the Department of Biostatistics and Medical Informatics include a network of 133 multi-core, 64-bit Linux servers (totaling more than 600 cores) and three Windows servers (with 20 cores). The facility currently houses 327 terabytes of enterprise-grade, networked storage configured in a redundant multi-homed setup. Most of these machines are made available for compute-bound tasks by a locally developed software system called Condor. Condor automatically locates workstations that are idle and transfers jobs to them. The jobs are periodically check pointed and migrate from machine to machine, as needed, until completion. Furthermore, the department has access to the campus-wide Condor Compute Cluster (GLOW) which provides an additional 2000+ CPUs. Additionally, access to the campus Condor High Throughput Computing (CHTC) facility can also be granted to specialized jobs. The CHTC facility contains 1900 computational cores running at least 2.8 GHz and 1.5GB of RAM per core. The Condor system provides excellent support for the extensive experimentation that is typical of machine-learning research. Additional resources available for this project include several Mac and Windows laptop computers.

A full complement of up-to-date computational tools are available, including Splus, R, SAS and Matlab for statistical exploration, as well as a number of optimizing compilers and a large suite of utilities. LaTex and Word are fully supported for producing publication-quality papers. For use in conjunction with the Gene Expression Center, the facility supports a dedicated server for Affymetrix LIMS and software packages.

The environment of the central facility is fully temperature-controlled and power conditioned and uses the Hospital's emergency power system. On a 24/7 basis, temperature, server functionality and security are also monitored by automated systems that notify an on-call staff member if problems arise. Remote access generally makes remote repairs possible during off-hours. All BCG-supported computers are connected to the Medical School's network and are behind its firewall, the maintenance of which are the responsibility of the BCG.

Highly experienced, full-time staff in the Department of Biostatistics and Medical Informatics provide such services as network access, file backup and recovery, software installation and support, maintenance of shared printers, etc.

SUPPORTING AND COLLABORATIVE RESOURCES

The Department of Biostatistics and Medical Informatics achieves vitality and strength through their relationship with a number of campus organizations including the UW Department of Statistics, the Biotech Center, and the UW Comprehensive Cancer Center (UWCCC.) Here we briefly review these critical resource components.

UW DEPARTMENT OF STATISTICS

The Department of Statistics and the Department of Biostatistics and Medical Informatics are closely related. Graduate students in the Department of Biostatistics and Medical Informatics may pursue an M.S. degree in Biostatistics, or a Ph.D. in Statistics with an emphasis in Biostatistics. Over 80% of the research faculty in the Department of Biostatistics and Medical Informatics have a partial appointment in the Department of Statistics.

The Department of Statistics, founded in 1960 by Professor George Box and residing in the College of Letters and Sciences, now boasts 27 faculty members, has in training approximately 100 graduate students, and awards 8-10 PhD degrees annually. From its inception, the Department evolved into one of the premier departments of statistics in the world. It has been ranked consistently among the top five departments in the country on various measures concerning quality of faculty and graduate programs. Today it reflects the breadth and diversity of theoretical and applied statistics. Some 403 PhD degrees and 563 MS degrees have been granted since 1963. Graduates pursue careers in academia, industry, or government. Roughly one third to one half of the PhDs work in biological or biomedical fields.

The quality and breadth of both research and teaching of the faculty are fundamental strengths of the Department. Ten of the current faculty members have attained the honor of election as Fellow of the American Statistical Association and eight that of Fellow of the Institute of Mathematical Statistics.

UW COMPREHENSIVE CANCER CENTER

The UW Comprehensive Cancer Center (UWCCC) is a multidisciplinary institution serving Wisconsin and adjoining areas of Illinois and Iowa. It conducts clinical and laboratory research on the biology of cancer, focused on human problems in etiology, prevention, tumor localization and treatment. Research programs include Immunology and Immunotherapy, Imaging and Radiation Sciences, Cancer Genetics, Cell Signaling and Growth Control, Etiology and Chemoprevention, Experimental Therapeutics, Human Cancer Virology, and Clinical Research Support across all programs. The UWCCC is among the world's leaders in research in breast, prostate and bladder cancers, medical onocolgy, immunobiology, radiation oncology, biostatistics, and medical physics.For the past 20 years the Biostatistics Shared Resource has been vital to the research life of the UWCCC. Faculty members from the Department of Biostatistics and Medical Informatics participate in this shared resource as consultants, collaborators, and trainers.

BIOTECHNOLOGY CENTER

The University of Wisconsin Biotechnology Center (UWBC), located at 425 Henry Mall, less than ¼ mile from the Statistics Department, offers state-of-the-art research services to UW-Madison scientists. These services include: DNA synthesis and sequencing, peptide synthesis, peptide sequencing and mass spectrometry of phosphopeptides and small metabolites, plant biotechnology, production of transgenic/knockout mice and rats, education programs and multimedia technology resources. The Center provides facilities for oligonucleotide synthesis and DNA sequencing, including two Illumina GA IIx instruments and two Illumina HiSeq2000 instruments for massive parallel sequencing (to be used in the ChIP-Seq experiments in this proposal). The Facility performs library preparation from fragmented DNA samples, maintains dedicated workstations for data analysis (using the CLC-Bio Genomics Workbench and other open source programs), and provides bioinformatic support and consultation services (including Bowtie mapping of reads). In addition, the Biotechnology Center performs probe preparation, hybridization, and scanning analysis of both Affymetrix and Nimblegen/Roche microarrays.

Other Resources

Library Access

The statistics journals and texts are housed in the Wendt Library located adjacent to the Statistics Department. Library facilities are available near both Biostatistics office spaces. The Clinical Science Center (CSC) houses Weston Medical Library, and the Medical Science Center (MSC) is adjacent to the Middleton Library. Thus, most relevant library resources are extremely convenient. Further, there is state-of-the art electronic access to a wide range of electronic journals and journal storage facilities. The UW Madison library system is one of the ten largest library systems in the country.

OMB Number: 4040-0001
Expiration Date: 06/30/2011

RESEARCH & RELATED Senior/Key Person Profile (Expanded)

PROFILE - Project Director/Principal Investigator

Prefix: | * First Name: Frederick | Middle Name: J
* Last Name: Boehm | Suffix:
Position/Title: POSTDOCTORAL TRAINEE | Department: Biostatistics/Med Info
Organization Name: The Board of Regents of the University of Wisconsin System | Division: SCHOOL OF MEDICINE AND PUBLIC
* Street1: 5730 Medical Sciences Center
Street2: 1300 University Ave
* City: Madison | County/ Parish: Dane
* State: WI: Wisconsin | Province:
* Country: USA: UNITED STATES | * Zip / Postal Code: 53706-1510
* Phone Number: 608-890-1248 | Fax Number:
* E-Mail: fjboehm@wisc.edu

Credential, e.g., agency login: fjboehm

*** Project Role:** PD/PI | **Other Project Role Category:**

Degree Type: M.D.
Degree Year: 2007

***Attach Biographical Sketch** BOEHMapplicant_fellowbiosketc [Add Attachment] [Delete Attachment] [View Attachment]
Attach Current & Pending Support [Add Attachment] [Delete Attachment] [View Attachment]

PROFILE - Senior/Key Person 1

Prefix: | * First Name: MICHAEL | Middle Name: A
* Last Name: NEWTON | Suffix:
Position/Title: PROFESSOR | Department: STATISTICS-GEN/BIOSTATISTICS
Organization Name: The Board of Regents of the University of Wisconsin System | Division: Medicine and Public Health
* Street1: 1300 UNIVERSITY AVE
Street2: 1245A MEDICAL SCIENCES CTR
* City: MADISON | County/ Parish:
* State: WI: Wisconsin | Province:
* Country: USA: UNITED STATES | * Zip / Postal Code: 53706-1509
* Phone Number: 608 263 0357 | Fax Number:
* E-Mail: NEWTON@BIOSTAT.WISC.EDU

Credential, e.g., agency login:

*** Project Role:** Other (Specify) | **Other Project Role Category:** Sponsor

Degree Type:
Degree Year:

***Attach Biographical Sketch** newton13biosketch1013448838.p [Add Attachment] [Delete Attachment] [View Attachment]
Attach Current & Pending Support [Add Attachment] [Delete Attachment] [View Attachment]

RESEARCH & RELATED Senior/Key Person Profile (Expanded)

PROFILE - Senior/Key Person 2

Prefix: | * First Name: MICHAEL | Middle Name: N

* Last Name: GOULD | Suffix:

Position/Title: Professor | Department: Oncology

Organization Name: The Board of Regents of the University of Wisconsin System | Division: Medicine and Public Health

* Street1: McArdle Lab for Cancer Research

Street2: 1400 University Avenue

* City: Madison | County/ Parish: Dane

* State: WI: Wisconsin | Province:

* Country: USA: UNITED STATES | * Zip / Postal Code: 53706-1509

* Phone Number: 608-263-6615 | Fax Number: 608-262-2824

* E-Mail: gould@oncology.wisc.edu

Credential, e.g., agency login:

*** Project Role:** Other (Specify) | **Other Project Role Category:** Co-Sponsor

Degree Type:

Degree Year:

***Attach Biographical Sketch**	gould_biosketch_v2_oldPersona	Add Attachment	Delete Attachment	View Attachment
Attach Current & Pending Support		Add Attachment	Delete Attachment	View Attachment

FELLOWSHIP APPLICANT BIOGRAPHICAL SKETCH

USE ONLY FOR INDIVIDUAL PREDOCTORAL and POSTDOCTORAL FELLOWSHIPS. DO NOT EXCEED FOUR PAGES.

NAME OF FELLOWSHIP APPLICANT	POSITION TITLE
Frederick Joseph Boehm	Postdoctoral Trainee
eRA COMMONS USER NAME (credential, e.g., agency login) fjboehm	

EDUCATION/TRAINING (Begin with baccalaureate or other initial professional education, such as nursing, and include postdoctoral training.)

INSTITUTION AND LOCATION	DEGREE *(if applicable)*	YEAR(s)	FIELD OF STUDY
University of Wisconsin-Madison, Madison, Wisconsin	B.S.	2001	Mathematics; Chemistry; Biochemistry
University of California-Berkeley, Berkeley, California	None		Organic chemistry
University of Wisconsin-Madison, Madison, Wisconsin	M.S.	2007	Population health sciences
University of Wisconsin-Madison, Madison, Wisconsin	M.D.	2007	Medicine
University of Washington, Seattle, Washington	None		Biostatistics
University of Wisconsin-Madison, Madison, Wisconsin	Ph.D. (anticipated)	2016	Statistics (with Biostatistics emphasis)

Please refer to the application instructions in order to complete sections A, B, C, and D of the Biographical Sketch.

A. Personal Statement

A desire both to understand biology and to reduce human suffering unifies my research and schooling over the last fifteen years. I have taken multiple approaches to achieve these noble goals. Recently I have realized the convergence of my medical, quantitative, and biological interests as my research has focused on biostatistical methods for the development of personalized medicine.

As an undergraduate student, I intensely studied multivalent cell-to-cell signaling from a chemist's point of view. I synthesized a series of carbohydrate-displaying polymers for use in cell binding assays. We described our findings in a 2002 manuscript. The University of Wisconsin-Madison recognized my contributions to this project by awarding me the Book Store Academic Excellence Award in 2001.

A few days after college graduation in 2001, I flew to Berkeley, California to begin graduate studies in organic chemistry. A serious illness in late 2001 disrupted my studies. After my recovery, I returned to school to study medicine and public health and graduated with degrees in medicine (M.D.) and population health sciences (M.S.) in 2007. Following graduation in 2007, I accepted a research position in the University of Washington's department of biostatistics. My work, in collaboration with NHGRI-funded GENEVA investigators, contributed to the development and dissemination of biostatistical methods for genome-wide association studies in humans. Then, in 2010, I returned to the University of Wisconsin-Madison to pursue a Ph.D. in statistics (with biostatistics emphasis). I formally enrolled in the statistics graduate program in 2011.

In my ambitious postdoctoral training program, I will draw on my quantitative, medical, and biological backgrounds to address the most urgent questions in human biology and medicine. Among these important tasks is the need for biostatistical tools that enable personalized genomic medicine. Personalized genomic medicine integrates a patient's biological and medical data, on an individual patient level, to inform medical decisions and to tailor therapies to each patient's clinical and biological profile. In the case of breast cancer, knowledge of a given tumor's genetic code will enable clinicians to administer therapies that are targeted specifically to the individual patient's unique tumor.

My formal schooling, when coupled with my recent training in biostatistical methods development, have uniquely positioned me to succeed in this rigorous training course and to develop an independent biomedical research career.

B. Positions and Honors

ACTIVITY/OCCUPATION	BEGINNING DATE (mm/yy)	ENDING DATE (mm/yy)	FIELD	INSTITUTION/COMPANY	SUPERVISOR/ EMPLOYER
Senior Fellow	11/07	09/09	Biostatistics	University of Washington	Bruce Weir
Assistant researcher	01/11	07/11	Econometrics	University of Wisconsin-Madison	Emin Dokumaci
Postdoctoral trainee	02/12		Biostatistics	University of Wisconsin-Madison	Michael Newton

Academic and Professional Honors

Howard Hughes Medical Institute Medical Student Research Training Fellowship, 2006 (declined due to security concerns in Haiti)
UW-Madison Medical School Shapiro Award for research with Partners In Health, 2004
UW-Madison LOCUS stipend for research with Partners In Health, 2004
National Science Foundation Graduate Research Fellowship, 2001
UW-Madison Undergraduate Excellence Award, 2001
University Book Store Award for Outstanding Undergraduate Thesis, 2001
UW Chemistry Department Summer Research Scholarship, Edward Panek Memorial Scholarship, 2000
Hilldale Undergraduate Research Fellowship, 1999
Phi Beta Kappa Honor Society, 1999 to present
Wisconsin Idea Undergraduate Fellowship, for Salvation Army Learning Center, with Anders Olson, 1999
Ralph B. Abrams Scholarship, 1999, awarded to UW-Madison College of Letters & Science students on the basis of merit
UW-Madison Chemistry Department Summer Research Scholarship, 1998
Margaret and Allard Smith Scholarship, 1998, awarded to UW-Madison College of Letters & Science students on the basis of merit
Dean's List, Spring 1997, Spring 1998, Fall 1998, Spring 1999, Fall 1999, Spring 2000, Fall 2000, Spring 2001
UW-Madison College of Letters & Science Honors Program 1996 to 2001
Wisconsin Academic Excellence Scholarship, 1996 to 2000

C. Publications

Peer-reviewed journal articles

1. “Quality control and quality assurance in genotypic data for genome-wide association studies", C.C. Laurie, et al. 2010, 34(6): 591.
2. “Economic risk factors for HIV infection among women in rural Haiti: Implications for HIV prevention policies and programs in resource-poor settings", M.C. Smith Fawzi, et al.. Journal of Women's Health 2010, 19(5):885.
3. “Contrasting identity-by-descent estimators, association studies, and linkage analyses using the Framingham Heart Study data", E.E. Marchani, et al.. BMC Proceedings 2009, 3(Suppl 7):S102.
4. “Cell aggregation by Scaffolded Receptor Clusters", J.E. Gestwicki, et al., Chemistry and Biology, 2002, 9:163.

Book chapters

1. “Urinalysis and an approach to kidney diseases", A.V. Moorthy and F.J. Boehm in Pathophysiology of kidney disease and hypertension, A.V. Moorthy, et al., eds. 2009. Elsevier. ISBN: 978-1-4160-4391-1.
2. “Acid-base homeostasis and metabolic alkalosis", K.A. Traeger, F.J. Boehm, and A. Djamali in Pathophysiology of kidney disease and hypertension, A.V. Moorthy, et al., eds. 2009. Elsevier. ISBN: 978-1-4160-4391-1.
3. “Metabolic acidosis and approach to acid-base disorders", F.J. Boehm, K.A. Traeger, and A. Djamali in Pathophysiology of kidney disease and hypertension, A.V. Moorthy, et al., eds. 2009. Elsevier. ISBN: 978-1-4160-4391-1.

Edited books

1. Pathophysiology of kidney disease and hypertension, A.V. Moorthy, B.N. Becker, F.J. Boehm, A. Djamali, eds. 2009. Elsevier. ISBN: 978-1-4160-4391-1.

Abstracts

1. “Synthesis of galactosyl-alpha-(1-3)galactosyl epitopes for use in immunological studies" E. Fasella, F.J. Boehm, and L. L. Kiessling. American Chemical Society meeting, San Francisco, California, March 2000. Abstracts of papers of the American Chemical Society 2000, Vol. 219, p. 97.

D. Scholastic Performance

YEAR	SCIENCE COURSE TITLE	GRADE	YEAR	OTHER COURSE TITLE	GRADE
	UW-Madison				
1996	Chemical Principles I	A	2005	Psychiatry	BC
1997	Chemical Principles II	A	2005	Primary Care	B
1997	Organic Chemistry I	A	2005	Surgery	B
1998	Organic Chemistry II	A	2006	Internal Medicine	BC
1998	Cell Biology	A	2006	Obstetrics & Gynecology	AB
1998	Biochemistry I	A	2006	Pediatrics	AB
1998	Organismal biology	A	2006	Obstetrics & Gynecology	AB

YEAR	SCIENCE COURSE TITLE	GRADE
1999	Biological Interactions	A
1999	Cellular Signal Transduction	A
1999	Analytical Chemistry	A
1999	Biochemistry II	A
1999	Inorganic Chemistry	A
1999	Biophysical chemistry	A
2000	Organic Chemistry III	A
2000	Physical chemistry	A
2000	Physical chemistry laboratory	A
2000	Biochemical Methods	A
2000	Spectrochemical measurements	A
2000	Symmetry, bonding, molecular shapes	A
2001	Protein-nucleic acid binding	A
2001	Topology	A
2001	Logic	A
2002	Introduction to biostatistics	A
2002	Epidemiology	A
2003	Epidemiological methods	AB
2003	Quantitative methods	A
2003	Epidemiology of infectious diseases	A
2003	Biochemistry	AB
2003	Histology	B
2003	Anatomy	B
2004	Medical genetics	AB
2004	Pathology	B
2004	Physiology	B
2004	Neuroscience	A
2004	Hematology	BC
2004	Cardiovascular System	C
2004	Neoplasia	A
2004	Infection and immunity	BC
2004	Pharmacology	B
2004	Respiratory System	B
2005	Renal	BC
2005	Nutrition	AB
2005	GI	B
2005	Endocrine	C
2005	Infection and immunity	B
2005	Psychiatry	B
2005	Pharmacology II	B
	UW Seattle	
2008	Biostatistics I	3.7
2009	Biostatistics II	4
2008	Statistical Inference I	2.9
2009	Statistical Inference II	3
2009	Linear Models	2.9
	UW Madison	
2011	Introduction to measure theory	A
2011	Mathematical statistics I	A
2011	Linear models	B
2011	Real analysis	C
2012	Mathematical statistics II	AB
2012	Design of Experiments	AB

YEAR	OTHER COURSE TITLE	GRADE
2006	Anesthesiology	AB
2006	Child psychiatry	AB
2006	General surgery	A
2006	Pediatrics Sub-internship	BC
2007	Neurosciences	B
2007	Preceptorship	A
2007	Medicine	AB
2007	Pathology	AB
2007	Radiology	AB
2007	Clinical chemistry	AB
2010	Biology of Microorganisms	B
2010	Chemical thermodynamics	B
2010	Quantum chemistry	B

UW-Madison grades on a scale of A(4.0), AB (3.5), B(3.0). UW-Seattle grades on a scale of 4.0, 3.9, 3.8, etc.

BIOGRAPHICAL SKETCH

Provide the following information for the Senior/key personnel and other significant contributors in the order listed for Form Page 2.
Follow this format for each person. **DO NOT EXCEED FOUR PAGES.**

NAME	POSITION TITLE
Newton, Michael A.	Professor
eRA COMMONS USER NAME manewton	

EDUCATION/TRAINING *(Begin with baccalaureate or other initial professional education, such as nursing, include postdoctoral training and residency training if applicable.)*

INSTITUTION AND LOCATION	DEGREE *(if applicable)*	MM/YY (s)	FIELD OF STUDY
Dalhousie University, Halifax, Nova Scotia	Diploma	1985	Engineering
Dalhousie University, Halifax, Nova Scotia	BSc	1986	Mathematics/Statistics
University of Washington, Seattle, WA	MS	1988	Statistics
University of Washington, Seattle, WA	PhD	1991	Statistics

A. Personal Statement

I am Professor in the Departments of Statistics and of Biostatistics and Medical Informatics. My research concerns statistical inference in various problems from the biological sciences, especially the use and development of stochastic models, empirical Bayesian methods, and advanced statistical computing. Examples include statistical methods for gene expression analysis, microRNA analysis, and the analysis of genomic aberrations. I work extensively in collaborative interdisciplinary projects, especially in genomics and cancer biology: new statistical methods are often required for the complicated data structures that are now routinely generated. In addition to research in statistical methodology, I am co-PI on a computationally-focused R01 project to identify functional properties of SNPs identified by breast cancer genome-wide association studies. I am PI on an HG project to develop new approaches to the integration of experimental and functional genomic data. I served on the Genome Study Section and was the founding biological sciences editor at the *Annals of Applied Statistics*. I co direct the Cancer Genetics Program at the UWCCC, I direct the Biostatistics Program at UW, and I am PI of the NIGMS funded T32, *Interdisciplinary Biostatistics Training Program*. My experience with genomic data analysis is extensive.

B. Positions and Honors.

1991-92	Assistant Professor, Statistics, University of Wisconsin at Madison (UW-Madison)
1992-96	Assistant Professor, Statistics and Biostatistics, UW-Madison
1992-present	Member, UW-Madison Comprehensive Cancer Center
1996-2002	Associate Professor, Statistics and Biostatistics & Medical Informatics, UW-Madison
2002-present	Professor, Statistics and Biostatistics & Medical Informatics, UW-Madison
2004-present	Director, Biostatistics Program, UW-Madison
2006-2009	Editor, Annals of Applied Statistics, Institute of Mathematical Statistics

Honors

1994	Young Investigator Award, Biometrics Section, American Statistical Association
1995-99	FIRST Award, National Cancer Institute
1997	George Snedecor Award, Committee of Presidents of Statistical Societies
2003	Mortimer Spiegelman Award, American Public Health Association
2004	Presidents' Award, Committee of Presidents of Statistical Societies
2007	Fellow, American Statistical Association
2011	Medallion Lecturer, Institute of Mathematical Statistics
2012-17	Kellett Mid-Career Researcher Award, UW Graduate School

C. Selected peer-reviewed publications related to genomic data analysis (Author of more than 80 papers since 1990; H-index 38)

Newton MA, Kendziorski CM, Richmond CR, Blattner FR, and Tsui KW. On differential variability of expression ratios: Improving statistical inference about gene expression changes from microarray data. *J Comp Bio* 8(1), 37-52, 2001.
Newton MA. Discovering combinations of genomic aberrations associated with cancer. *J Amer Statist Assoc* 97, 931-42, 2002.
Kendziorski CM, Newton MA, Lan H, and Gould MN. On parametric empirical Bayes methods for comparing multiple groups using replicated gene expression profiles. *Statist Med,* 22, 3899-3914, 2003.
Newton MA, Noueiry A, Sarkar D, and Ahlquist P. Detecting differential gene expression with a semi-parametric hierarchical mixture method. *Biostatistics* 5, 155-176, 2004.
Thliveris, AT, Halberg, RB, Clipson, L, Dove, WF, Sullivan, R, Washington, MK, Stanhope, S, and Newton, MA (2005). Polyclonality of familial murine adenomas: Analyses of mouse chimeras with low tumor multiplicity suggest short-range interactions. *Proc. Natl. Acad. Sci. USA*, 102, 6960-6965
Newton MA, Quintana FA, den Boon JA, Sengupta S Ahlquist P. Random-set methods identify distinct aspects of the enrichment signal in gene-set analysis. *Annals of Applied Statistics, 1, 85-106,* 2007.
Pyeon D, Newton MA, Lambert PF, den Boon JA, Sengupta S, Marsit CJ, et al. Fundamental differences in cell cycle deregulation in human papillomavirus–positive and human papillomavirus–Negative Head/Neck and Cervical Cancers. *Cancer Research* 67, 4705-4619, 2007.
Hao L, Sakurai A, Watanabe T, Sorensen E, Nidom CA, Newton MA, Ahlquist P, Kawaoka Y (2008). Drospophila RNAi screen identifies host genes important for influenza virus replication. *Nature*, 454, 890-893. PMID 18615016. PMCID: PMC2574945.
Stanhope, SA, Sengupta, S, den Boon, J, Ahlquist, P, and Newton, MA. (2009). Statistical use of Argonaute expression and RISC assembly in microRNA target identification. *PLoS Computational Biology*, 5(9): e1000516. PMCID: PMC2739424.
Newton, MA and Chung, LM (2010). Gamma-based clustering via ordered means with application to gene expression analysis. *Annals of Statistics*, 38, 3217-3244. PMCID: PMC2990889
Teague, B, Waterman MS, others and Schwartz, DC (2010). High resolution human genome structure by single-molecule analysis. *PNAS,* 107, 10848-10853. PMID 20534489. PMCID: PMC2890719.
Taapken, SM, Nisler, BS, Newton, MA, Sampsell-Barron, TL, Leonhard, KA, McIntire, EM, and Montgomery, KD (2011). Karyotypic abnormalities in human induced pluripotent stem cells and embryonic stem cells. *Nature Biotechnology*, 29, 313-314.
Amos-Landgraf, JM, Irving, AA, Hartman, C, Hunter, A, Laube, B, Chen, X, Clipson, L, Newton, MA, and Dove, WF (2012). Monoallelic silencing and haploinsufficiency in early murine neoplasms. *Proc. Natl. Acad. Sci.* 109, 2060-2065.
Newton, MA, He, Q, and Kendziorski, C (2012). A model-based analysis to infer the functional content of a gene list. *Statistical applications in genetics and molecular biology.* Vol 11, Issue 2, Article 9.
Sarkar, D, Goldstein, S, Schwartz, DC, and Newton, MA. Statistical significance of optical map alignments. *Journal of Computational Biology* , 19(5): 1-15. PMC3342520

D. Research Support. List selected ongoing or completed (during the last three years) research projects (federal & non-federal support).

Ongoing

R21HG006568 2/15/12-1/31/14
NIH
Computational statistics for model-based functional genomic data integration
The goal is to assess the feasibility of a function-centered approach to combining experimental and functional data to allow genomic scientists to better interpret the content of their experimental data. Role: PI

P30 CA14520 George Wilding (PI) 01/01/76-03/31/13
National Institutes of Health /National Cancer Institute
UW Comprehensive Cancer Center Support
Support for senior and program leaders of cancer center; administration and evaluation of cancer research and cancer center members; support of shared resources and services for peer-reviewed, cancer-related projects; developmental support for new investigators, projects and shared projects. Dr. Newton is senior statistician responsible primarily for laboratory collaborations and projects in the Cancer Center.
Role: Biostatistician

R01 ES017400 (Newton/Gould) 12/11/08 -10/31/13
NIH
Breast Cancer GWAS: Function and Environmental Interactions
The goal of this project is to develop an integrated approach combining global genetic information together with environmental exposure to form a network model that begins to describe the etiology of breast cancer
Role: Co-PI

1R01 CA123438 (Halberg) 4/1/09-2/28/14
NIH
Polyclonal Intestinal Tumors: formation, Progression, and Significance
The goal is to answer fundamental questions about polyclonality using a unique combination of newly developed mouse models, statistical techniques and imaging platform.
Role: Statistician, co-investigator

P01 CA 022443 (Sugden) 2/1/97-4/30/13
NIH
Molecular Biology and Genetics of Human Tumor Viruses
The goal of Project 1 is to further define and understand the viral functions and virus:host interactions that mediate HPV-associated disease from the point of infection to development of HPV-associated malignancy.
Role: Statistician, co-investigator

R37-CA63677 William Dove (PI) 03/01/03-6/30/13
NIH
Analyses of Progression to Colon Cancer in a Spectrum of Pathways
Dr. Newton is the project statistician working on experimental designs, mapping of quantitative factors affecting tumor susceptibility in animal models, and methodology for handling complex phenotypes. Role: Statistician, co-investigator

R21CA170876 Halberg (PI) 09/01/12-8/31/14
NIH
Molecular Differences Predicting Tumor Progression in Colorectal Cancer
We plan to monitor tumors as they progress, collectingbiopsies for histopathological assessment and molecular analysis. Transcriptional changes associated with progression can then be elucidated with DNA microarrays. The profile of adenomas that progress to invasive adenocarcinomas will be compared to the profile of adenomas that do not progress. Role: Co-Investigator

R01 MH061285 Pollak (PI) 09/01/12-8/31/14
NIH
Emotion Processing: Risk for Psychopathology in Children
The goal is to examine how brain circuitry is shaped by environmental experience in ways tat lead to childhood metal health problems using novel neuroscience-informed methods to measure the efficiency of emotional learning. Role: Co-Investigator

R01HG000225 Schwartz (PI) 1/1/91-6/30/15
NIH
New Physical Methodologies for Genomic Analysis
The goal is to enable new technologies that will provide sufficient physical map information to intimately mixz with modern sequencing data for comprehensive assembly of complex genomes. Role: Co-Investigator

<u>COMPLETED</u>

5 R01DE015944 Elaine Smith (PI) 4/1/04-3/31/09
National Institutes of Health
Transcripitonal and Genetic Profiles in HNSCCs
This consortium project integrates DNA microarrays, bioinformatics and epidemiology to identify signature patterns of gene expression in head and neck squamous cell carcinomas (HNSSCs), to advance

understanding of oncogenesis and to discover novel biomarkers for diagnosing HNSCC, for developing more effective, case-specific treatment, and for predicting survival.
Role: Biostatistician

R01 CA64364 Michael Newton (PI) 3/6/04-2/28/09
National Institutes of Health
Statistical Methods for Molecular Cancer Data
As PI, Dr. Newton is responsible for developing statistical methodology appropriate to the analysis and interpretation of allelic-imbalance data and related data on molecular and cytogenetic abnormalities.
Role: Principal Investigator

R21HG004379 (Schwartz) 8/1/07-7/31/10
NIH
Sequence Acquisition from Mapped Single DNA Molecules
The goal is to develop a scheme for acquisition of sequence information using large genomic DNA molecules that will be optically barcoded and analyzed for sequence content. Role: Statistician, co-investigator

17-2008-1043(Hoffman) 9/1/08-8/31/10
Juvenile Diabetes Research Foundation
New TGF-beta Signaling Inhibitors for Diabetic Nephropathy
The project goal is to have one or more optimized lead compounds identified in two years and available in sufficient quantities for testing in animal models of diabetic nephropathy. Role: Statistician, co-investigator

11/30/12

Program Director/Principal Investigator (Last, First, Middle): Boehm, Frederick J.

BIOGRAPHICAL SKETCH

Provide the following information for the Senior/key personnel and other significant contributors in the order listed on Form Page 2.
Follow this format for each person. **DO NOT EXCEED FOUR PAGES.**

NAME	POSITION TITLE
Michael N. Gould	Kelly Clifton Professor of Oncology
eRA COMMONS USER NAME (credential, e.g., agency login) MNGOULD	

EDUCATION/TRAINING *(Begin with baccalaureate or other initial professional education, such as nursing, include postdoctoral training and residency training if applicable.)*

INSTITUTION AND LOCATION	DEGREE *(if applicable)*	MM/YY	FIELD OF STUDY
University of Wisconsin, Madison, WI	B.S.	1969	Zoology
University of Wisconsin, Madison, WI	M.S.	1973	Radiological sciences
University of Wisconsin, Madison, WI	Ph.D.	1977	Radiological sciences
Argonne National Laboratory, Argonne, IL	Postdoc	1977-78	

Please refer to the application instructions in order to complete sections A, B, C, and D of the Biographical Sketch.

A. Personal Statement

The focus of my research, beginning in Graduate School, is on the etiology and prevention of breast cancer. The goal of the proposed research is to investigate the function of a non-coding breast cancer susceptibility locus that is found in both women and rats. I have broad experience in breast cancer research and most recently in the area of inherited genetics underlying breast cancer risk and breast cancer prevention. I have published over 200 papers with approximately 150 focusing on cancer research. Many of the proposed experiments use rat carcinogenesis models including genetically engineered rats of which I have extensive experience with. For example, our lab produced the first targeted knockout rat. Our research philosophy has been one of following a problem to its conclusion using the most appropriate technologies and areas of study as needed. Often this lands us in new areas of research.

In this proposal we incorporate novel research in several areas including immunology. I have been working in this field for the last four years and feel comfortable pursuing aims in this field. I have chosen an excellent immunologist to work with on this project. Jenny Gumperz, an innate T-cell immunologist, and I have been collaborating on developing this project over the last full year. In summary, I am very excited by the science and translational potential of this project and am fully committed to intensely pursue it.

B. Positions and Honors

Instructor, Depts. of Human Oncology and Radiology, Univ. of Wisconsin, Madison, WI, 1978-1979
Assistant Professor, Departments of Human Oncology and Radiology, University of Wisconsin, Madison, WI, 1979-1983
Visiting Scientist, Radiation Effects Research Foundation (Atomic Bomb CasualtyCommission), Hiroshima, Japan, 1982, 1984
Associate Professor, Departments of Human Oncology and Medical Physics, andEnvironmental Toxicology Program,1983-1988
Member, University of Wisconsin Comprehensive Cancer Center, 1978-present
Professor, Department of Human Oncology, University of Wisconsin, Madison, WI, 1988-1998
Professor, Department of Medical Physics, University of Wisconsin, Madison, WI, 1988-present
Director, Breast Cancer Program, University of Wisconsin Comprehensive Cancer Center, 1994-2001
Professor, Department of Oncology, University of Wisconsin, Madison, WI, 1998-present
Kelly H. Clifton Professor, University of Wisconsin, Madison, WI, 2000-present
Director, Etiology and Chemoprevention Program, UWCCC, 2001-2005
Oncology Boundary Study Section Review Committee – NIH, 2001
NIH NCI Merit Award, 1996-2005

NCI Board of Scientific Counselors, 2004-2009
Member, Genomics Center, University of Wisconsin, Madison, WI, 2007-present
Director, Cancer Genetics Program, University of Wisconsin Comprehensive Cancer Center, University of Wisconsin, Madison, WI, 2007-present
Interagency Breast Cancer and Environment Coordinating Committee (Research Process Subcommittee Chair) 2010-present

C. Selected Peer-Reviewed Publications (total publications >190)

Papers Specific for Mcs5a

1. Smits, B. M. G., Traun, B. D., Devries, T. L., Tran, A., Samuelson, D. J., Haag, J. D., and Gould, M. N. An Insulator Loop Resides Between the Synthetically Interacting Elements of the Human/Rat Conserved Breast Cancer Susceptibility Locus MCS5A/Mcs5a. Nucleic Acids Res., 40: 132-147, 2012. PMCID: PMC3245909
2. Smits, B. M. G., Sharma, D., Samuelson, D. J., Woditschka, S., Mau, B., Haag, J. D., and Gould, M. N. The Non-Protein Coding Breast Cancer Susceptibility Locus Mcs5a Acts in a Non-Mammary Cell-Autonomous Fashion Through the Immune System and Modulates T-Cell Homeostasis and Functions. Breast Cancer Res., 13:R81, 2011. PMCID: PMC3236344
3. Gould, M. N. The Utility of Comparative Genetics to Inform Breast Cancer Prevention Strategies. Genetics, 183: 409-412, 2009. PMCID: PMC2766305
4. , D. J., Hesselson, S. E., Aperavich, B. A., Zan, Y., Haag, J. D., Trentham-Dietz, A., Hampton, J. M., Mau, B., Chen, K.-S., Baynes, C., Khaw, K.-T., Luben, R., Perkins, B., Shah, M., Pharoah, P. D., Dunning, A. M., Easton, D. F., Ponder, B. A., and Gould, M. N. Rat Mcs5als a Compound Quantitative Trait Locus with Orthologous Human Loci That Associate with Breast Cancer Risk. Proc. Natl. Acad. Sci. USA, 104: 6299-6304, 2007.
5. , D. J., Aperavich, B. A., Haag, J. D., and Gould, M. N. Fine Mapping Reveals Multiple Loci and a Possible Epistatic Interaction within the Mammary Carcinoma Susceptibility Quantitative Trait Locus, Mcs5. Cancer Res., 65: 9637-9642, 2005.
6. Drinkwater, N. R., and Gould, M. N. The Long Path from QTL to Gene. PLoS Genetics, in press, 2012. PMCID: PMC Journal – In Process

Rat Breast Cancer Models

6. Cotroneo, M. S., Haag, J. D., Zan, Y., Lopez, C. C., Thuwajit, P., Petukhova, G. V., Camerini-Otero, R. D., Gendron-Fitzpatrick, A., Griep, A. E., Murphy, C. J., Dubielzig, R. R., and Gould, M. N. Characterizing a Rat Brca2 Knockout Model. Oncogene, 26: 1626-1635, 2007.
7. Woditschka, S., Haag, J. D., Waller, J. L., Monson, D. M., Hitt, A. A., Brose, H. L., Hu, R., Zheng, Y., Watson, P. A., Kim, K., Lindstrom, M. J., Mau, B., Steele, V. E., Lubet, R. A., and Gould, M. N. Neu-Induced Retroviral Rat Mammary Carcinogenesis: A Novel Chemoprevention Model for Both Hormonally Responsive and Nonresponsive Mammary Carcinomas. Cancer Res., 66: 6884-6891, 2006.
8. Zan, Y., Haag, J. D., Chen, K.-S., Shepel, L. A., Wigington, D., Wang, Y.-R., Hu, R., Lopez-Guajardo, C. C., Brose, H. L., Porter, K. I., Leonard, R. A., Hitt, A. A., Schommer, S. L., Elegbede, A. F., and Gould, M. N. Production of Knockout Rats using ENU Mutagenesis and a Yeast-based Screening Assay. Nat. Biotechnol., 21: 645-651, 2003.
9. Watson, P. A., Kim, K., Chen, K.-S., and Gould, M. N. Androgen-dependent Mammary Carcinogenesis in Rats Transgenic for the Neu Proto-Oncogene. Cancer Cell, 2: 67-79, 2002.

Selected Relevant Papers (chronological order)

10. Veillet, A. L., Haag, J. D., Remfert, J. L., Meilahn, A. L., Samuelson, D. J., and Gould, M. N. Mcs5c: A Mammary Carcinoma Susceptibility Locus Located in a Gene Desert that Associates with Tenascin C Expression. Cancer Prev. Res., 4: 97-106, 2011. NIHMSID: NIHMS304756
11. Cotroneo, M. S., Merry, G. M., Haag, J. D., Lan, H., Shepel, L. A., and Gould, M. N. The Mcs7 Quantitative Trait Locus Is Associated with an Increased Susceptibility to Mammary Cancer in Congenic Rats and an Allele-Specific Imbalance. Oncogene, 25: 5011-5017, 2006.
12. Berchtold, C. M., Chen, K.-S., Miyamoto, S., and Gould, M. N. Perillyl Alcohol Inhibits a Calcium-Dependent Constitutive Nuclear Factor-**k**B Pathway. Cancer Res., 65: 8558-8566, 2005.
13. Kamiya, K., Yasukawa-Barnes, J., Mitchen, J. M., Gould, M. N., and Clifton, K. H. Evidence that Carcinogenesis Involves an Imbalance Between Epigenetic High Frequency Initia¬tion and

Suppression of Promotion. Proc. Natl. Acad. Sci. USA, 92: 1332-1336, 1995.
14. Hsu, L.-C., Kennan, W. S., Shepel, L. A., Jacob, H. J., Szpirer, C., Szpirer, J., Lander, E. S., and Gould, M. N. Genetic Identification of Mcs-1, a Rat Mammary Carcinoma Suppressor Gene. Cancer Res., 54: 2765-2770, 1994.
15. Gould, M. N., Biel, W. F., and Clifton, K. H. Morphological and Quantitative Studies of Gland Formation from Inocula of Monodispersed Rat Mammary Cells. Exp. Cell Res., 107: 405 416, 1977.

D. Research Support

ONGOING

R01 ES017400 (Gould, M., Newton, M. co-PIs) 12/11/08 to 10/31/13

Breast Cancer GWAS: Function and Environmental Interactions

The aims of this proposal are: (1) to integrate expression quantitative trait loci data with several genome-wide association studies for breast cancer risk using high throughput gene expression measurements from at least 50 reduction mammoplasty HMEC samples from women without breast disease, along with full genome panels of single nucleotide polymorphism genotypes; and (2) to further refine the functional networks developed in Aim 1 by investigating the effects of environmental factors.

R01 CA077494 (Gould, PI) 6/1/09 to 5/31/13

Wky Rat Genetic Model for Breast Cancer Susceptibility

The aims of this project are: (1) to extend genetic and biological studies using existing congenic recombi nant rat models for the Mcs5cand Mcs5bloci; and (2) to annotate the reduced genomic intervals of Mcs5cand Mcs5bby comparing the Wky and WF alleles in order to identify welldefined candidate elements that underlie the susceptibility phenotypes of these loci.

U01 ES019466 (Gould, PI) 9/1/10 to 4/30/15

Genetics of Breast Cancer Risk at Windows

The overall goal of this multi-investigator project is to quantify and characterize three post-natal windows of susceptibility to breast cancer using rat mammary carcinogenesis models and then to confirm the results in human studies.

PENDING

11233158 (Gould, PI) 11/1/12 to 10/31/13

Epigenomic Adaptation to Low Dose Radiation

PHS Fellowship Supplemental Form

OMB Number: 0925-0002

A. Application Type:

From SF424 (R&R) Cover Page. The response provided on that page, regarding the type of application being submitted, is repeated here for your reference as you provide the responses that are appropriate for this Fellowship application.

☒ New ☐ Resubmission ☐ Renewal ☐ Continuation ☐ Revision

B. Research Training Plan

1. Introduction to Application *(for RESUBMISSION applications only)*		Add Attachment	Delete Attachment	View Attachment
2. * Specific Aims	2_SpecificAims_v221013457483.pdf	Add Attachment	Delete Attachment	View Attachment
3. * Research Strategy	3_researchStrategy_v33_NOREFS10134	Add Attachment	Delete Attachment	View Attachment
4. Inclusion Enrollment Report *(for RENEWAL applications only)*		Add Attachment	Delete Attachment	View Attachment
5. Progress Report Publication List *(for RENEWAL applications only)*		Add Attachment	Delete Attachment	View Attachment

Human Subjects

Please note. The following item is taken from the Research & Related Other Project Information form. The response provided on that page, regarding the involvement of human subjects, is repeated here for your reference as you provide related responses for this Fellowship application. If you wish to change the answer to the item shown below, please do so on the Research & Related Other Project Information form; you will not be able to edit the response here.

Are Human Subjects Involved? ☐ Yes ☒ No

6. Human Subjects Involvement Indefinite? ☐ Yes ☐ No

7. Clinical Trial? ☐ Yes ☐ No

8. Agency-Defined Phase III Clinical Trial? ☐ Yes ☐ No

9. Protection of Human Subjects		Add Attachment	Delete Attachment	View Attachment
10. Inclusion of Women and Minorities		Add Attachment	Delete Attachment	View Attachment
11. Targeted/Planned Enrollment		Add Attachment	Delete Attachment	View Attachment
12. Inclusion of Children		Add Attachment	Delete Attachment	View Attachment

Other Research Training Plan Sections

Please note. The following item is taken from the Research & Related Other Project Information form. The response provided on that page, regarding the use of vertebrate animals, is repeated here for your reference as you provide related responses for this Fellowship application. If you wish to change the answer to the item shown below, please do so on the Research & Related Other Project Information form; you will not be able to edit the response here.

Are Vertebrate Animals Used? ☐ Yes ☒ No

13. Vertebrate Animals Use Indefinite? ☐ Yes ☐ No

14. Vertebrate Animals		Add Attachment	Delete Attachment	View Attachment
15. Select Agent Research		Add Attachment	Delete Attachment	View Attachment
16. Resource Sharing Plan		Add Attachment	Delete Attachment	View Attachment
17. * Respective Contributions	17_RespectiveContributions_v310134	Add Attachment	Delete Attachment	View Attachment
18. * Selection of Sponsor and Institution	18_SelectionOfSponsorAndInstitutio	Add Attachment	Delete Attachment	View Attachment
19. * Responsible Conduct of Research	19_ResponsibleConductOfResearch_v4	Add Attachment	Delete Attachment	View Attachment

PHS Fellowship Supplemental Form

C. Additional Information

Human Embryonic Stem Cells

1. * Does the proposed project involve human embryonic stem cells? ☐ Yes ☒ No

If the proposed project involves human embryonic stem cells, list below the registration number of the specific cell line(s) from the following list: http://stemcells.nih.gov/research/registry/. Or, if a specific stem cell line cannot be referenced at this time, please check the box indicating that one from the registry will be used:

☐ Specific stem cell line cannot be referenced at this time. One from the registry will be used.

Cell Line(s):

Fellowship Applicant

2. Alternate Phone Number:

3. Degree Sought During Proposed Award:

Degree:	If "other", please indicate degree type:	Expected Completion Date (month/year):	
PHD: Doctor of Philosophy		05/2016	Reset Entry

4. * Field of Training for Current Proposal: 4410 Biostatistics and/or Biometry

5. * Current Or Prior Kirschstein-NRSA Support? ☒ Yes ☐ No

If yes, please identify current and prior Kirschstein-NRSA support below:

* Level	* Type	Start Date *(if known)*	End Date *(if known)*	Grant Number *(if known)*	
Postdoctoral	Institutional	02/01/2012	01/31/2013	T32ES007015	Reset Entry
					Reset Entry
					Reset Entry
					Reset Entry

6. * Applications for Concurrent Support? ☐ Yes ☒ No

If yes, please describe in an attached file: Add Attachment | Delete Attachment | View Attachment

7. * Goals for Fellowship Training and Career: C7_GoalsForFellowshipTrainingAnd — Add Attachment | Delete Attachment | View Attachment

8. * Activities Planned Under This Award: C8_activitiesPlannedUnderThisAwa — Add Attachment | Delete Attachment | View Attachment

9. Doctoral Dissertation and Other Research Experience: C9_ResearchExperience_v510134574 — Add Attachment | Delete Attachment | View Attachment

10. * Citizenship:
☒ U.S. Citizen or noncitizen national
☐ Permanent Resident of U.S. Pending
☐ Permanent Resident of U.S.
(If a permanent resident of the U.S., a notarized statement must be provided by the time of award)
☐ Non-U.S. Citizen with temporary U.S. visa

Tracking Number:GRANT11287005 Funding Opportunity Number:PA-11-113 Received Date:2012-12-10T11:28:47-04:00

PHS Fellowship Supplemental Form

C. Additional Information (continued)

Institution

11. ☐ Change of Sponsoring Institution

Name of Former Institution:

D. Sponsor(s) and Co-Sponsor(s)

* Sponsor(s) and Co-Sponsor(s) Information: II_sponsorInfoWithMNLetter_v101013 | Add Attachment | Delete Attachment | View Attachment

E. Budget

All Fellowship Applicants:

1. * Tuition and Fees:

☐ None Requested ☒ Funds Requested:

Year 1	30,579.00
Year 2	31,496.00
Year 3	
Year 4	
Year 5	
Year 6 *(when applicable)*	
Total Funds Requested:	62,075.00

Senior Fellowship Applicants Only:

2. Present Institutional Base Salary: Amount | Academic Period | Number of Months | Reset Entry

3. Stipends/Salary During First Year of Proposed Fellowship:

a. Federal Stipend Requested: Amount | Number of Months

b. Supplementation from other sources: Amount | Number of Months

Type (sabbatical leave, salary, etc.)

Source

F. Appendix

Add Attachments | Delete Attachments | View Attachments

Tracking Number:GRANT11287005 Funding Opportunity Number:PA-11-113 Received Date:2012-12-10T11:28:47-04:00

Specific aims

Personalized genomic medicine (PGM), in which an individual patient's clinical and genomic profile guides or informs medical decision-making, will allow physicians and patients to tailor therapies to the individual patient's molecular disease sources. The promise of PGM requires methods and tools 1) to pinpoint molecular interactions and aberrations that cause disease and, subsequently, 2) to identify diseased molecules and pathways in individual patients. In initial progress towards PGM, scientists have invested tremendous resources in genome-wide association studies (GWAS) in the last decade. Proper integration of GWAS findings with other genomic data will bridge the chasm between GWAS findings and PGM implementation. Tools for PGM ultimately will lead to higher breast cancer detection rates and a more diverse arsenal of breast cancer therapies.

A natural extension of GWAS are studies that address whether disease-associated SNPs impact gene expression in relevant tissues. We use the term “eQTL”, an abbreviation of “expression quantitative trait locus”, to refer to SNPs whose genotypic variation associates with difference in mean expression level of one or more transcripts. While many investigators have undertaken eQTL studies in post-GWAS SNP characterization, high dimensionality of genomic data complicates eQTL study design. Sample size planning for eQTL studies remains an open question. Our approach extends mixture model-based sample size estimation methods from two-class gene microarray designs to the eQTL study design setting. We will:

Aim 1: Develop methods to estimate eQTL study sample size with false discovery rate control in the setting of post-GWAS SNP characterization.

Challenge: To model high dimensionality and multiple measured and unmeasured sources of variation in expression data while accounting for not merely two classes, but all three SNP genotype classes.

Approach: We will adapt and extend existing hierarchical mixture model-based methods that 1) control false discovery rate and 2) permit more than two classes of subjects. We will examine these methods via simulation and in our pilot genomics data.

Impact: These methods will enable eQTL studies for efficient post-GWAS SNP characterization in breast cancer and, ultimately, other human diseases. Furthermore, eQTL studies serve as a critical link in the path that connects GWAS to clinical medicine. Better eQTL detection methods will accelerate development of personalized genomic medicine.

Aim 2: Develop sample size planning methods for pathway-based inference in eQTL studies.

Challenge: To extend mixture model methods to accommodate sample size calculation for biological pathway-based eQTL studies.

Approach: We will develop flexible hierarchical mixture models to allow for pathway-based inference. We then will assess our methods via simulation studies and in our pilot genomics data.

Impact: Our approach will enhance understanding of non-coding SNP biology by allowing for detection of biological pathways, rather than single genes, whose transcription levels differ by SNP genotype class.

Completion of these aims will bridge a gap between GWAS results and their clinical applicability through personalized genomic medicine. Furthermore, the proposed studies represent an outstanding opportunity for the fellowship applicant to develop as an independent biomedical scientist.

RESEARCH STRATEGY

A. SIGNIFICANCE

Importance: understanding biological mechanisms that underlie disease-associated genomic loci

Personalized genomic medicine requires molecular-level understanding of biological mechanisms by which genetic variants impact phenotypes. Personalized genomic medicine's tailoring of therapies and preventative measures to individual patients will aid not only individual patients, but will enhance health for the public. Better therapies for breast cancer will lead to fewer premature deaths and, thus, improve public health.

Necessary tools for developing personalized genomic medicine include a detailed understanding of biomolecular interactions and biological pathways. Fortunately, genomic technologies have enabled progress towards this understanding. The NHGRI catalog of published GWAS findings contains 7,925 phenotype-associated SNPs[1]. While scientists have invested tremendous energy and resources in GWAS, many opportunities exist to extend these results towards the goal of achieving personalized genomic cancer therapies[2]. Most GWAS discover disease-associated SNPs that aren't in exons[2]. Recent research has supported the hypothesis that SNP genotypes associate with disease status by impacting gene expression in relevant regions of disease-implicated tissues[3–5]. These disease-associated SNPs in non-coding regions may impact gene expression levels by, for example, altering transcription factor binding[6].

Critical barriers: to biostatistical methods and tools for sample size planning in eQTL studies for post-GWAS SNP characterization

While scientists have conducted more than 1400 reported GWAS, these have, to date, yielded relatively few clinical and medical improvements[1,2]. To reap the full societal and public health benefits of these studies, we need to take the next steps towards personalized genomic medicine. Our proposed research enables these goals.

Furthermore, biostatistical software tools that implement methods for post-GWAS SNP characterization are few in number. As we describe below, development and dissemination of our new methods and software tools will enable biological study of cancer-associated genomic loci and provide valuable insights into molecular mechanisms in a wide range of diseases.

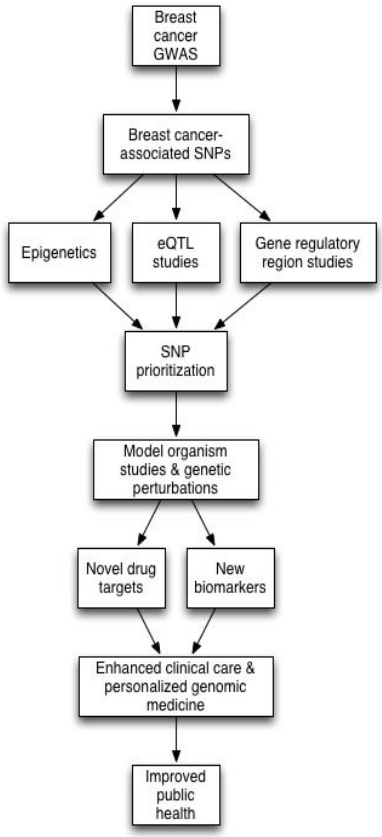

Figure 1. Path from GWAS to eQTL studies to personalized genomic medicine and improved public health.

Improvement of scientific knowledge: understanding molecular disease mechanisms and biological networks

Our project to develop methods for eQTL sample size planning promises to aid in understanding biological pathways and disease mechanisms[3]. While we address these questions in the context of breast cancer, our methods will provide tools for studies that will lead to insights in other human diseases. Our project thus serves as a case study in how to develop quantitative methods and tools to study biology.

Improvement of clinical practice: through novel biomarkers, new drug targets, and refined preventative measures

Understanding biology of disease-associated SNPs ultimately will enable improvements in clinical care[7]. In the case of breast cancer, a more complete biological understanding of disease processes will provide novel drug targets and new biomarkers to aid in early diagnosis and disease monitoring.

B. APPROACH

Overview of the team and the approach

At the heart of our project are 1) a highly productive, two-decade research collaboration between world-class scholars Michael Gould and Michael Newton and 2) the training needs for developing multi-disciplinary scholar Frederick Boehm. Newton is an expert in applying multivariate statistics to biological questions. Gould has dedicated his professional career to understanding breast cancer genetics. Statistics graduate student Boehm will benefit from a synergy of Newton's and Gould's respective strengths. Under their close mentorship, Boehm is uniquely situated to accomplish his research goals, to build on his training in medicine, mathematics, chemistry, statistics, and biology, and to develop into an independent biomedical scientist with the capacity to undertake groundbreaking statistical genomics research.

The approach to the two research aims requires Boehm to develop expertise in both multivariate statistics and breast cancer molecular biology. Given his prior medical training and experience with high-dimensional GWAS data, he is in an excellent position to accomplish these goals in a timely fashion. With direction from Newton and Gould, Boehm will develop these skills while extending statistical methods to eQTL studies. His research foci will be 1) mixture model-based methods development for sample size planning in post-GWAS eQTL studies in breast cancer and 2) extending mixture models to aid in sample size planning for pathway-based inference. He will use our pilot genomics data from 62 women to inform his simulations and to assess his mixture model-based methods.

Software development, dissemination, resource sharing, and reproducible research

Michael Newton has extensive experience with software development and dissemination. He is co-developer of the “EBArrays” Bioconductor package. Although it was first developed more than six years ago, users have downloaded “EBArrays” more than 2500 times in the last 12 months. With Newton's guidance, Boehm will write a user-friendly, open-source R package for the proposed methods and share it through the Bioconductor repository.

With a great need for reproducibility in statistical research, we will prepare literate programming documents using Sweave or knitr to enable full reproduction of our analyses[8,9]. We will share these documents and our accompanying tutorial on the Bioconductor website[10]. Furthermore, we will submit our R code to “Run My Code” for validation and companion website creation[11].

Timeline

Work on Aim 1 will begin immediately and will lead to manuscript #1. With contributions from Gould and Newton, Boehm will write manuscript #1 to describe methodological research involving sample size planning in eQTL studies. We will submit manuscript #1 to a scientific journal for publication before the end of fellowship year 1. Work on Aim 2 will begin in the second half of the first fellowship year, after Boehm has expanded his statistical knowledge base through coursework in Bayesian methods. At that point Boehm will have completed nearly all of the statistics Ph.D. coursework. Boehm will write, with assistance from Gould and Newton, manuscript #2 to describe simulation results from aim 2 studies. We will use manuscript #3 to describe integration of our novel methods to the pilot genomics data on 62 healthy women. We will submit manuscripts #2 and #3 before January 1, 2015, which is 18 months after the proposed fellowship start date. During the final 6 months of the fellowship period, Boehm will prepare an independent research proposal (possibly for K mechanism

funding through NIH) to continue and to expand his biostatistical methods research. This will serve as the culmination of Boehm's fellowship training in scientific writing. It also would fund Boehm's first years as an independent biomedical scientist.

Preliminary studies supporting the approach

We describe below our collaborative research that has inspired the current proposal.

Pilot study in post-GWAS SNP characterization

Boehm, Newton, and Gould have collaborated since Boehm's entry into the statistics graduate program in August 2011. During that time, we worked together to perform a pilot study involving SNP and expression array data from 62 healthy women. We collected reduction mammoplasty breast tissue samples from 80 women without history of a cancer diagnosis. We extracted DNA for SNP genotyping on Illumina's Omni chip. We extracted RNA for Affymetrix 1.0 GeneChip analysis. After removing low-quality samples, we worked with data from 62 women. We implemented further quality control measures for genotype data[12]. During this process, we discovered and corrected a sample mislabeling. We then proceeded with initial eQTL studies.

In our initial eQTL analyses, we limited consideration to SNP loci that associated with breast cancer status in GWAS[1]. We used several statistical approaches, including ANOVA methods and two-sample t-test methods, to detect trans and cis eQTL. Weak differential expression patterns in our data complicated our analyses and biological inferences. We persevered with our studies and developed software tools for post-GWAS SNP characterization, which we are still refining before submission to Bioconductor's software repository. Our proposed strategy, with its focus on sample size planning for post-GWAS eQTL studies, results from concerns over the weak differential expression patterns in our pilot genomic data.

Open-source software for reproducible research in genomics

We conducted our analyses in the R statistical computing environment[13]. To enable reproducibility, we have prepared Sweave documents that combine into a single tutorial document 1) our analysis code and 2) annotated methods descriptions. We will freely share our open-source code and accompanying tutorial via the Bioconductor software repository[10].

We now re-state aim 1, contextualize it, and discuss in detail our approach to accomplishing it, along with potential obstacles and alternative approaches. We subsequently do the same for aim 2.

Aim 1: Develop methods for eQTL study sample size planning in post-GWAS SNP characterization.

To enable eQTL identification among breast cancer GWAS-implicated SNPs, and to bridge the gap between GWAS results and advances in clinical medicine, we will develop methods to plan eQTL study sample size. We now frame our scientific question in its genomics context and provide necessary background for mixture models and false discovery rates before describing our plan.

Background on post-GWAS eQTL studies

Post-GWAS eQTL studies in breast cancer, leukemia, colorectal cancer, kidney disease, and other human diseases have made progress towards functionally annotating phenotype-associated GWAS SNPs[14]. Despite these initial successes, a thorough understanding of SNP biology requires further characterization. Adapting the reasoning of Freedman and co-workers, we envision a three-pronged approach (Figure 1) to post-GWAS SNP biology in breast cancer: 1) eQTL studies, 2) epigenetics studies, and 3) gene regulatory region studies (including transcription factor binding sites)[2]. Together, these three approaches enable SNP prioritization to inform model organism studies and, ultimately, to enable personalized genomic medicine and improved public health.

Researchers have used a wide range of sample sizes in post-GWAS eQTL studies, from 60 subjects in a study of cis-only eQTL among chronic lymphocytic leukemia-associated SNPs to more than 1900 subjects in a post-GWAS cis- and trans-eQTL study of type I diabetes[15,16]. However, there are few sample size estimation methods that account for high-dimensionality of expression data in the eQTL setting.

Background on hierarchical mixture models

We begin by explaining our use of the term “hierarchical mixture model”[17]. The description of observed variation by 1) latent random variables and 2) the conditional variation given realizations of the latent random variables is the feature that distinguishes a statistical model as being “hierarchical”. Latent random variables in our model include 1) gene-specific expected values and 2) SNP genotype class-specific expected values. Such a hierarchical model makes it possible to “borrow” information from across genes to make gene-specific inferences[18].

We use the term “mixture model” to refer to our use of gene-specific hypotheses about differential expression as latent discrete random variables[17]. In the case of only two classes, each gene is either differentially expressed (DE) or equivalently expressed (EE) between the two classes. We then think of each gene as a Bernoulli trial, or coin flip, where the two outcomes are DE and EE.

We now explain our hierarchical mixture model for three classes, which is needed for modeling of three SNP genotype classes. By expanding the number of components in the mixture, we accommodate a greater number of differential expression patterns. For instance, in study of SNP genotype classes, we want to stratify our subjects into three classes (AA, AB, and BB), according to genotype at a single SNP. In this case, we need to consider five expression patterns: one pattern of equivalent expression in all three groups, and four patterns of differential expression (three in which two groups are equivalent and one in which all three classes are different). We denote the (weighted) equivalent expression pattern as $p_0 f(AA_j, AB_j, BB_j)$, and the four (weighted) differential expression patterns as: $p_1 f(AA_j) f(AB_j, BB_j)$, $p_2 f(AB_j) f(AA_j, BB_j)$, $p_3 f(BB_j) f(AA_j, AB_j)$, and $p_4 f(AA_j) f(AB_j) f(BB_j)$. Note that equivalently expressed classes share a joint density function, while differentially expressed classes have distinct marginal densities. The p_i are the weights, i.e., the proportion of genes with the i^{th} expression pattern. For example, p_3 is the proportion of genes for which genotype classes AA and AB are equivalently expressed, while class BB expression differs from that of AA and AB, as might be seen in a situation where A has a dominant effect on gene j transcription, while B is recessive. Summing these five terms, we obtain the marginal distribution of gene-level data:

$$p(AA_j, AB_j, BB_j) = p_0 f(AA_j, AB_j, BB_j) + p_1 f(AA_j) f(AB_j, BB_j) + p_2 f(AB_j) f(AA_j, BB_j) + p_3 f(BB_j) f(AA_j, AB_j) + p_4 f(AA_j) f(AB_j) f(BB_j)$$

The five expression patterns correspond to hypotheses about possible clustering of mean expression levels across all N subjects. We now let $d_j = (d_{j,1}, d_{j,2}, \ldots, d_{j,N})$ be the vector of all N observations of gene j. We mix over the five above patterns of DE and EE to get the joint distribution:

$$p(d_j) = p_0 f_0(d_j) + p_1 f_1(d_j) + p_2 f_2(d_j) + p_3 f_3(d_j) + p_4 f_4(d_j)$$

We complete hierarchical model specification by recognizing that each of the five terms above is itself a product of densities from each clustering component.

$$f_k(d_j) = \prod_{i=1}^{r(k)} f(d_{j,S_{k,i}}) = \prod_{i=1}^{r(k)} \int \left(\prod_{s \in S_{k,i}} f_{obs}(d_{j,s} | \mu) \pi(\mu) d\mu \right)$$

where S is the set of the first N positive integers, $r(k)$ is the number of subsets in the partition

corresponding to the k^{th} expression pattern, where k is an integer from 0 to 4, and $S_{k,i}$ is the i^{th} subset of the k^{th} expression pattern's partition of S. The random effects distribution that determines the expected values of the latent, gene-specific expression levels appears above as $\pi(\mu)$. The observed component of the hierarchical model is denoted f_{obs}. We then use the data to estimate 1) weights, p_i, 2) parameters for f_{obs}, and 3) parameters for mean component $\pi(\mu)$.

Background on false discovery rate methods

Storey adapted Benjamini and Hochberg's false discovery rate (FDR) methods to the gene expression setting[19–21]. We consider m hypothesis tests, where m is a fixed positive integer. For instance, we might have m genes in an expression study. Using notation from Jorstad and colleagues, we have a table of outcomes (Table 1)[22]. Storey proved the following relation for positive false discovery rate (pFDR)[20].

$$pFDR(\alpha)=\frac{\pi_0\alpha}{Pr(p<\alpha)}=Pr(H_0\,true|H_0\,rejected)$$

Jorstad and colleagues and Hu et al. note that this equation connects three concepts: 1) pFDR, 2) significance region (through α), and, through $Pr(p<\alpha)$, 3) sample size N[22,23]. They note that, with the above equation for pFDR, one can estimate sample size N after appropriate choices of pFDR and α [22].

Three-class sample size planning with FDR control in eQTL studies

	H_0 accepted	H_0 rejected	Total
H_0 true	A_0	R_0	m_0
H_0 false	A_1	R_1	m_1
Total	M_A	M_R	m

Table 1. Outcomes of m hypothesis tests. All quantities except m are random variables.

We will fit the above hierarchical mixture model using standard expectation-maximization (EM) methods. We then will use parameter estimates from model fits and reasonable values of α and pFDR to estimate $Pr(p<\alpha)$. Our estimate of $Pr(p<\alpha)$, in turn, will yield the sample size estimate.

Methods assessment

We will use simulation studies and pilot genomics data from 62 reduction mammoplasty tissue specimens to assess our novel sample size estimation methods.

Potential obstacles & alternative approaches

One concern is the potential need to consider multiple computational algorithms for model fitting. In this case, we will draw on Michael Newton's exceptional experience in statistical computing to identify alternative, efficient algorithms for model fitting.

Aim 2: Develop sample size planning methods for pathway-based inference.

A goal in biological pathway-based inference is to detect transcription patterns among functionally related genes. Intrinsic flexibility of our hierarchical mixture model framework will accommodate pathway summary statistic inputs rather than single transcript-level inputs.

Background for pathway-level inference in eQTL studies

Pathway-level inference provides a complementary perspective to trancript-level methods. One strength of pathway-level inference is that it enables detection of regulated pathways in which few or none of the member genes meet statistical significance thresholds in single-gene analyses. On the other hand, a potential limitation of pathway-level methods is a reliance on annotation; that is, methods may rely on presence of annotated pathway databases, such as Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), or Reactome to group transcripts by functional similarity and may not recognize previously unannotated biological pathways[26–28].

Our approach to sample size planning methods for pathway-based inference

Instead of using transcript-level inputs as in aim 1, we will fit our hierarchical mixture model with pathway summary inputs. Newton and co-workers have developed methods for random gene set methods[29,30]. Newton has implemented these methods and compatibility with KEGG and GO in the R software package allez[31]. We will use pathway-level summary statistics, such as those from Newton and co-workers' random gene set methods, as inputs to a hierarchical mixture model29. We will then follow the logic in aim 1 for eQTL sample size planning.

Methods assessment

We will need to consider a variety of pathway-level summary statistics when evaluating our methods. We want to understand how different strategies to account for pathway characteristics, such as number of genes in a pathway, impact sample size planning.

Potential obstacles and alternative approaches

In addition to potential model fitting obstacles, we may need to expand our consideration of pathway-level summary statistics beyond those resulting from random set methods. We will study other pathway-level summary statistics as needed.

Respective contributions

Collaborative process among Boehm, Newton, and Gould in development, review, and editing of research training plan

Research training plan development involved close collaboration among Frederick Boehm, Michael Newton, and Michael Gould. We communicated via email and in person. Gould's vision guided our research focus from preliminary studies of pilot genomic data towards sample size assessment in post-GWAS SNP characterization studies. Newton's extensive contributions to mixture model methods in gene expression studies informed our approach. Newton articulated the statistical interplay among mixture models, false discovery rate methodology, and related empirical Bayes methods that is central to the proposed research.

Boehm wrote a first draft of the research training plan. Newton and Gould read and critiqued it. Newton's suggestions focused on how to relate elements of statistical theory to questions of biological importance. Gould's comments focused on emphasizing biological implications of the proposed research. Newton's previous manuscripts that describe hierarchical mixture models and their application in gene expression settings heavily influenced the proposed plan. Boehm and Newton extensively edited subsequent training plan drafts with inputs from Gould.

Discussion of respective roles in accomplishing proposed research

Overview of collaborative interactions

Gould and Newton will guide Boehm as he uses the proposed plan to develop into an independent biomedical scientist. Through 1) weekly in-person discussions with Boehm and 2) review of Boehm's monthly progress reports they will provide critical feedback and supervision of Boehm's training experience. Should the project encounter obstacles to progress, Boehm has immediate access to Gould and Newton to discuss and overcome obstacles.

Newton will meet weekly with Boehm to discuss progress towards developing the proposed biostatistical methods. Boehm will spend one half-day per week in the Gould laboratories. During this time, Boehm will meet with Gould each week to discuss research progress and to learn from Gould's experimental biology findings. Furthermore, Boehm will meet with one or more of Gould's other trainees during the weekly laboratory visit. This meeting with trainee Bart Smits or another colleague will update Boehm on progress in breast cancer research, from a biologist's perspective, and give Boehm an opportunity to discuss biostatistical methods with biologists.

Aim 1 roles

Newton will guide Boehm in his study of hierarchical mixture models, false discovery rates, and related empirical Bayes methods. Much of Newton's research in the last decade has addressed these topics, which makes Newton the ideal mentor for Boehm's study. Once Boehm has developed sufficient familiarity with relevant biostatistical literature, he will perform, with guidance from Newton and the statistical literature, sample size planning in eQTL studies. Boehm will prepare and share computer code to implement his methods. Gould's major role will be to provide critical resources, including pilot genomic data, and to aid Boehm in understanding biological implications of their research findings.

Aim 2 roles

Both Newton and Gould will provide critical feedback on Boehm's implementation of pathway-level summary statistics. Gould will again provide pilot genomics data and biological context for the findings. Newton's experience with random gene set methods will allow Newton to guide Boehm's approaches to summarizing evidence of gene regulation across known pathways.

Selection of sponsor, co-sponsor, and institution

“We believe that the great state University of Wisconsin should ever encourage that continual and fearless sifting and winnowing by which alone the truth can be found.” This statement, from an 1894 Board of Regents report, explains why I have chosen the University of Wisconsin-Madison as my training institution. I can imagine no intellectual community that is more dedicated to the courageous quest for scholarly understanding.

While rule of thumb dictates that one study at distinct institutions for undergraduate and graduate work, I feel that my change of disciplines from college to graduate school exempts me from this rule. In college, I studied in the departments of chemistry, biochemistry, and mathematics. I now seek to train in a distinct set of disciplines: biostatistics, statistics, oncology, and genomics. I will learn and develop new methods in an intellectual atmosphere that is new to me and that I am still creating through my interactions with scholars and trainees across campus. Furthermore, I will use the professional relationships that I developed during college and medical school to enhance my training through scientific collaborations. Thus, rather than being an apparent weakness in my application, my decision to train at UW-Madison is very much a strength of my proposal.

I am compelled to train with Newton and Gould because of their statuses as world-class scholars, their excellent training records, their demonstrated support of me and my development as a scholar, their history of productive collaborations, and the vibrant intellectual training atmosphere that they and their coworkers have created for me. Both Newton and Gould have heavily invested in me and my research interests.

Newton is an internationally recognized leader in biostatistical methods development. While many scholars point to Stanford University's Bradley Efron as the leading statistician, Efron's work regularly cites and praises Newton's research. From our first meeting, I have admired Newton's passion and desire to understand statistical questions. I consider it a tremendous privilege to work with him.

Gould's career at the McArdle Cancer Research Center has focused on understanding breast cancer etiology. He has developed numerous rodent models for breast cancer. He and his trainees have made major contributions to molecular cancer genetics, including discovery and characterization of mammary cancer susceptibility genetic loci. Gould's unwavering optimism and dogged determination to understand biology inspire me every day. His critical role in this project is one of contextualizing our statistical methods and results into their biological setting. Newton's and Gould's investment in and dedication to my development as an independent biomedical scientist compel me to train with them.

Responsible conduct of research training plan

Boehm's RCR plan will enable him to engage research ethics in a scholarly fashion. The plan has four components: 1) IRES fellowship application, 2) seminar participation, 3) coursework in research ethics, and 4) a capstone project in research ethics. These experiences extend Boehm's previous RCR training, which includes enrollment in a 15-week course on research ethics during the Fall 2012 semester and enrollment in a 10-week course in ethics in stastistics from April to June 2009.

1. IRES fellowship application. Boehm will apply for the UW-Madison's Integrating Research Ethics and Scholarship (IRES) fellowship program (http://www.grad.wisc.edu/ethics/iresfellows.html). He will submit his application for admission in June 2013. If admitted to the competitive program's small cohort, he will participate fully in the IRES fellowship program during the 2013-2014 academic year. Due the financial support of the NIH, Boehm will not receive the usual $3000 stipend for the IRES fellowship program. The IRES fellowship program has not yet released its 2013-2014 calendar. 2012-2013 fellowship events included: attend full-day retreat on August 28, 2012, meet weekly from 8:30-9:30 a.m. Wednesday, September 12- Wednesday November 14 (ten sessions), attend full-day retreat on Thursday, January 17, 2013, attend two spring brown bags presented by 4 other fellows (these meetings start on 8:30am to 10:30am Wednesday, January 23 and run until February 20, present a brown bag utilizing the activity you will use for your large event, create and implement a research ethics event for your department, create an ethics portfolio in which you reflect on your knowledge and ability to teach research ethics.

2. Seminar participation. UW-Madison's IRES program also hosts seminars (that are open to all students) during the academic year. Boehm will attend all six of these per academic year for the entirety of his training period. Upcoming and past seminars include: Interactive Movie Night: The Lab, Tuesday, October 16, 2012, 4:00-5:30pm; The Ethics of Social Media Research, Tuesday, October 23, 2012, 4:00-5:30pm; and Collaborations with Statisticians, Wednesday, December 5, 2012, 4:00-5:00pm. In addition to IRES seminars, Boehm will attend medical history and bioethics department seminars. Past and upcoming seminars of interest include: “Escaping Melodrama: The US STD Inoculation Studies in Guatemala and the Difficulties for an Historian”, 4pm December 3, 2012 and “Water Water Everywhere: Human Experiments in Drinking Seawater,” 4pm November 17, 2011.
To learn more about recent scholarship in research ethics and to engage local scholars in this field, Boehm will attend annual bioethics symposiums in April 2013 and April 2014 (http://www.med.wisc.edu/event/fourth-annual-bioethics-symposium/26544).

3. Coursework in research ethics. Boehm will enroll in “Medical history and bioethics 545: Ethical and regulatory issues in clinical investigation” in September 2013. Twelve weekly meetings for 2 hours each will allow Boehm to “explore and examine the ethical issues central to clinical research, regulations governing clinical investigation, and the role of good clinical practice for clinical trials” and “to think critically about the ethical issues central to clinical research and know the basic elements of the federal regulations affecting clinical investigation.” (http://medhist.wisc.edu/courses/listcourses.php?year=2012&term=fall).

4. Capstone project in research ethics. One of Boehm's major interests within RCR is the nature of collaborations between statisticians and non-statisticians. This interest will be the focus of his capstone project. He will enroll in statistics 998 “Statistical consulting” in the fall 2014 term. An old syllabus is in the appendix. This course addresses in detail the nature of collaborations, and will provide some of the relevant background for his capstone project. The heart of his capstone project consists of 1) preparation of a written literature review on statisticians' and non-statisticians' attitudes about collaborations and 2) a December 2014 public seminar that summarizes his literature review. With the goal of informing his seminar presentation, he will informally discuss collaborations between statisticians and other scientists with his colleagues in toxicology, statistics, and biostatistics. Boehm will devote at least 3 hours per week to the literature review and seminar preparation during the 15-week fall 2014 term. Michael Newton and Michael Gould will critique the literature review and seminar slides.

Goals for fellowship training and career

I will use this training experience to develop into an independent scholar in biostatistics. My future independent research will address urgent biomedical questions in personalized genomic medicine and statistical genomics. My ongoing training in biostatistics and my previous medical training have positioned me to excel in this field. My medical training provided me with an understanding of human biology and disease pathology. My biostatistics training integrates this medical knowledge with my quantitative research interests. Furthermore, I will use the fellowship years not only to develop biomedical research skills, but also to learn, under the guidance of world-class scientists, professional skills that will enable me to be a fully independent biomedical scientist. Professional skills that I will hone during my fellowship include those related to grant proposal writing, manuscript writing, public speaking, and research mentoring.

Skills

I am eager to continue learning biomedical research and professional development skills from two exceptional mentors, Michael Newton and Michael Gould. Much instruction will be through writing. In addition to meeting at least once per week with Michael Newton and Michael Gould, I will prepare monthly written progress reports for Professors Gould and Newton. They will read and will discuss the reports with me. Through this mechanism, I will learn to formulate and to articulate scientific research questions and findings. Furthermore, through written reports and verbal and written feedback, I will learn how to develop testable hypotheses from my scientific questions and how to design experimental and analytical approaches to test my hypotheses. The monthly written reports will allow me the opportunity to get detailed, constructive feedback on my interpretation of results and my use of preliminary results to inform subsequent experiments and analyses. Furthermore, Newton and Gould have eagerly agreed to aid me as I develop writing skills for a wider audience through research manuscript preparation and manuscript submission to scientific journals. Preparation and discussions of the monthly reports will aid me as I progress to writing full research manuscripts.

Theories

I will learn elements of advanced mathematical statistics theory primarily from Michael Newton. Professor Newton will advise me as I learn 1) multivariate statistics, 2) mixture models, 3) false discovery rate theory, and 4) empirical Bayes theory. This instruction will be through independent readings and weekly discussions with Professor Newton. Furthermore, I will enroll in courses with Professor Newton and Professor Chunming Zhang to formalize my studies of areas 1) and 4), respectively. I will address areas 2) and 3) through reading monographs and primary literature, under the guidance of Professor Newton. Michael Newton has expertise in all four of the above areas of statistics, and is the perfect person to guide me through the mathematics and statistical theory that must be understood before developing novel methodology and statistical theory in these fields of biostatistics.

Conceptual approaches

Newton and Gould will guide me as I develop an intellectual framework for applying quantitative methods and tools to biomedical research questions. Within this framework, we'll consider the question of how to enable public health benefits from recent advances in both genomic technologies and biostatistical methods.

Activities planned under this award

Boehm will devote time to research, coursework, seminars, responsible conduct of research training, professional development activities, teaching, and research mentoring. Coursework is detailed in the sponsor and co-sponsor information section and includes a sufficient series of classes both to fulfill degree requirements for the Ph.D. in statistics with biostatistics option and to encourage Boehm's interest in multivariate biostatistics. Boehm plans to attend at least four seminars per week in statistics, biostatistics, genomics, and oncology during the academic years of his fellowship period. Furthermore, he will participate in formal coursework and seminar opportunities in responsible conduct of research. Further details of his RCR training plan are described in the appropriate section.

Year	Activity	Percentage time
1	Research	50%
1	Coursework	30%
1	Seminars	10%
1	Responsible conduct of research training	5%
1	Professional development	3%
1	Teaching	2%
2	Research	50%
2	Coursework	20%
2	Seminars	10%
2	Teaching	5%
2	Research mentoring	5%
2	Professional development	5%
2	Responsible conduct of research training	5%

Boehm's professional development activities have four components: 1) one-on-one meetings with Michael Gould and Michael Newton, 2) UW-Madison graduate school professional development workshops, 3) UW-Madison writing center workshops and individualized writing instruction, and 4) formal coursework in professional development skills. He plans to use these resources to develop skills such as grant proposal writing, public speaking, and manuscript writing that will enable him to become an independent researcher. Boehm also plans to teach a summer short course in statistical computing in July 2013 and July 2014. This partially fulfills his personal obligation to disseminate statistical methods.

Beginning in the second fellowship year, Boehm will devote less time to coursework and, consequently, will have more time to spend with professional development activities and teaching. He also looks forward to his first research mentoring experience in the second fellowship year. He will simultaneously mentor an undergraduate researcher in biostatistics and participate in a course that addresses research mentoring through the UW-Madison's scientific teaching program.

Research experience

Dates: September 1998 to May 2001
Discipline: Chemical biology
Supervisor: Professor Laura Kiessling
Project: Ring-opening metathesis (ROM) polymers for cell binding assays.

We synthesized, in large scale, mannose-bearing ROM polymers of lengths ranging from 10 to 1000 monomer units. We then used these molecules to study multivalency in cell adhesion molecules. We described our findings in a 2002 manuscript ("Cell aggregation by scaffolded receptor clusters", Gestwicki et al., Chemistry & Biology 9(2): 163-9). I also described a portion of this project in my undergraduate thesis, for which the UW-Madison gave me its undergraduate thesis award. Our study demonstrated that our multivalent mannose-bearing polymers enabled cell aggregation when cell-surface concanavalin A (ConA) was present. In the absence of ConA, little cell aggregation was observed. Furthermore, we observed that longer polymers led to greater extents of cell adhesion. We thus concluded 1) that we had devised a scale-able strategy for synthesizing chemical biology tools to study cell adhesion, 2) that our polymers enabled cell adhesion via conA, and 3) that increased valencies lead to enhanced degrees of cell adhesion.

Dates: May 2004 to August 2004
Discipline: Infectious disease epidemiology
Supervisor: Mary Kay Fawzi, Sc.D.
Project: Risk factors for HIV infection in women in rural Haiti.

We analyzed case-control data from a medical center in rural Haiti with the aim of identifying social and economic risk factors for HIV infection among this highly vulnerable population. Our co-workers administered detailed surveys to gather demographic, social, and economic data from participants. We reported findings from our multivariate logistic regression analyses in a 2010 manuscript ("Economic risk factors for HIV infection among women in rural Haiti: implications for HIV prevention policies and program in resource-poor settings", Fawzi et al., J. Women's Health 19(5): 885-92) and a 2011 presentation ("Economic risk factors for syphilis infection among pregnant women in rural Haiti", J. Mark et al., Sexually Transmitted Infections 87(1): 263). We found that a number of culturally sensitive survey question responses associated with HIV status. Our anthropologist colleagues indicated that the associated responses related to degree of poverty. This led us to conclude that, even in a poor community such as that in rural Haiti, there is economic stratification within the community, and the very poorest are at greatest HIV risk. This research has implications for distribution of resources in extremely poor communities.

Dates: November 2007 to September 2009
Discipline: Statistical genomics
Supervisor: Professor Bruce Weir
Project: Quality control and quality assurance for genotypic data in genome-wide association studies.

We collaborated with the NHGRI-funded GENEVA investigators to conduct a series of genome-wide association studies. Our specific responsibilities included developing and implementing quality control and quality assurance measures for genotypic data. We published our methods and findings in 2010 ("Quality control and quality assurance in genotypic data for genome-wide association studies", Laurie et al., Genetic Epidemiology 34(6): 591-602). We identified several important quality control measures, including examination of B-allele frequency (BAF) and log R ratio (LRR) plots for SNPs on each chromosome. This work has inspired other studies that relate BAF and LRR plot findings to cancer prevalences. Our conclusions from this study include a recognition of the importance of quality control and quality assurance in high-throughput genotyping technology.

Dates: August 2011 to present
Discipline: Statistical genomics
Supervisors: Professor Michael Newton and Professor Michael Gould
Project: Expression QTL detection among breast cancer-associated SNPs in healthy subjects.

We investigated a number of statistical methods, including ANOVA and two-sample t-test approaches, for detecting associations between SNP genotypes and expression levels. We faced many challenges in our study. High-dimensionality of our data (approximately 25,000 expression traits and approximately 1 million SNPs) and our small sample size of 62 subjects contributed to our difficulties. We now will investigate methods of sample size planning in post-GWAS eQTL studies.

SECTION II – SPONSOR AND CO-SPONSOR INFORMATION

II.A: Research support available

Faculty name	Role & Grant title	Source, number, status	Project period	Current year direct costs awarded
Newton	PI: Computational statistics for model-based functional genomic data integration	NIH; 1R21HG006568; Active	02/15/12-01/31/14	**$125,000**
Newton	Co-investigator: UWCCC Cancer Center Support Grant; PI: Wilding	NIH; 5P30CA014520-39; Active; renewal pending	01/01/76-03/31/13	**$3,309,051**
Newton & Gould	Co-PIs: Breast Cancer GWAS: Function and Environmental Interactions	NIH; 5 R01ES017400-04; Active	12/11/08-10/31/13	**$279,669**
Newton	Co-Investigator: Polyclonal Intestinal Tumors: Formation, Progression, and Significance	NIH; 5R01CA123438-04; Active	04/01/09-02/28/14	**$201,275**
Newton	Co-Investigator: Analyses of Progression to Colon Cancer in a Spectrum of Pathways; PI: Dove	NIH; 4 R01CA063677-17; Active	09/22/93-06/30/13	**$242,500**
Newton	Co-Investigator: Molecular Biology and Genetics of Human Tumor Viruses; PI: Sugden	NIH; P01CA022443-35; Active	02/01/97-04/30/13	**$1,284,297**
Newton	Co-Investigator: Molecular Differences Predicting Tumor Progression in Colorectal Cancer; PI: Halberg	NIH; 1R21CA170876-01; Active	09/01/12-8/31/14	**$130,500**
Newton	Co-Investigator: Emotion Processing: Risk for Psychopathology in Children	NIH; R01MH061285-11; Active	04/01/00-05/31/17	**$381,999**
Newton	Co-Investigator: New Physical Methodologies for Genomic Analysis	NIH; 2R01HG000225-16; Active	01/01/91-6/30/15	**$474,800**
Newton	PI: Interdisciplinary Biostatistics Training Program	NIH; T32GM074904-08 Active	7/11/05-6/30/13	**$170,128**
Gould	PI: Wky Rat Genetic Model for Breast Cancer Susceptibility	NIH; R01 CA077494 Active	6/1/09-5/31/13	**$210,000**
Gould	PI: Genetics of Breast Cancer Risk at Windows of Exposure	NIH; U01 ES019466 Active	9/1/10-4/30/15	**$297,000**
Gould	PI: Epigenomic Adaptation to Low Dose Radiation	DOE; 11233158; Pending	11/1/12 to 10/31/13	**$175,000 requested**

II.B. Sponsor's and co-sponsor's previous trainees

Current or past	Trainee name	Supervisor	Pre- or post-doctoral	Training period	Project title	Current position or current support
Past	Hoff, P.	Newton	Pre	1994-2000	Constrained Nonparametric Estimation Via Mixtures	Associate Professor of Statistics at University of Washington
Past	Dahl, D.	Newton	Pre	1999-2004	Conjugate Dirichlet Process Mixture Models: Efficient Sampling, Gene Expression, & Clustering	Associate Professor of Statistics at Texas A&M
Past	Yang, H.	Newton	Pre	1999-2005	Model-Based Clustering of Genomic Aberrations: Generalizing the Instability Selection Network Model	Assistant Professor of Biostatistics at Duke
Past	Chung, L.	Newton	Pre	2004-2010	Statistical Methods for the Analysis of Gene Expression Data	Department of Biostatistics, Yale University
Current	Pei, Q.	Newton	Pre	2006-	Statistical Models and SNP Detection Methods for Flash Sequencing	RA funded by RO1 (D. Schwartz)
Past	Samuelson, D.	Gould	Post	2001-2007	Mammary cancer genomics	Assistant Professor, U. of Louisville
Past	Zan, Y.	Gould	Post	1999-2003	Rat genomics	Associate researcher, UW-Madison
Current	Wiederholt, C.	Gould	Post	2005-	Breast cancer susceptibility	NIH research funds
Current	Smits, B.	Gould	Post	2006-	Molecular biology of breast cancer resistance loci	DOD Era of Hope Postdoctoral Fellowship
Current	Fautch, J.	Gould	Post	2009-	Breast cancer susceptibility	NIH research funds

II.C.1. Training plan Boehm is simultaneously a post-doctoral researcher and a student in the statistics Ph.D. program with biostatistics emphasis. He entered the statistics Ph.D. program in August 2011, and is eagerly preparing for the August 2013 statistics department qualifying exam. His strong undergraduate experience in basic sciences, such as chemistry, mathematics, and biology, forms a solid foundation for a career in biomedical research. He now seeks to integrate his quantitative, medical, and biological interests into a career as an academic biostatistician in translational cancer research. His desire to develop biostatistical methods for personalized genomic medicine is the perfect interdisciplinary blend for a young scientist with his interests and abilities.

II.C.1.i. Classes. Boehm plans to take, at a minimum, the courses in Table 3. He may revise this list as course schedules are announced. The goals of the coursework are both to accelerate Boehm's training in formal statistics and to enhance his professional development. Such training is needed to enable Boehm to develop independence as a biostatistician.

Academic term	Course department & number	Course title
Spring 2013	Statistics 710	Mathematical Statistics
Spring 2013	Biostatistics and Medical Informatics 826	Personalized Genomic Medicine
Spring 2013	Statistics 992	Large Scale Inference
Spring 2013	Botany 575	Scientific writing
Fall 2013	Statistics 641	Statistical Methods in Clinical Trials
Fall 2013	Computer sciences 760	Machine learning
Fall 2013	Statistics 775	Bayesian statistics
Fall 2013	Medical history and bioethics 545	Ethical and regulatory issues in clinical investigation
Spring 2014	Statistics 642	Statistical Methods in Epidemiology
Spring 2014	Biostatistics and Medical Informatics 776	Advanced Bioinformatics
Spring 2014	Oncology 675	Appropriate conduct in science
Spring 2014	Statistics 741	Survival analysis
Fall 2014	Statistics 998	Statistical consulting
Fall 2014	Statistics 760	Multivariate statistics
Fall 2014		
Fall 2014		

Table 3. Courses from January 2013 to December 2014.

II.C.1.ii. Seminars. Boehm plans to attend weekly seminars during all academic terms of his fellowship (Table 4). He also plans to attend monthly human genetics journal club meetings and seminars in the McArdle Cancer Research Center, the Wisconsin Institutes of Discovery, the Crow Institute for the study of evolution, and the departments of genetics and medical genetics.

II.C.1.iii. Opportunities for interactions with other groups and scientists. Since February 2012, a NIEHS training grant has supported Boehm's research. As part of this training program, he has attended weekly toxicology seminars. He has presented his research at this seminar and at the annual training grant retreat. Furthermore, he has used these opportunities to network with scientists in other disciplines.

Additionally, since February 2011, Boehm has attended weekly meetings of the Kendziorski research team to discuss issues in statistical genomics. He has shared much of his experience with GWAS with the Kendziorski team as he has learned about their diverse research interests. The spring 2013 presents a special opportunity for Boehm, as

Seminar title	Day	Time
Statistics Department Seminar	Wednesday	4pm
Biostatistics and Medical Informatics Department Seminar	Friday	12pm
Computation and Informatics in Biology and Medicine Seminar	Tuesday	4pm
Molecular and Environmental Toxicology Center Seminar	Thursday	3pm
Statistics Department Student Seminar	Wednesday	5:30pm
Kendziorski Research Group Meeting	Monday	1:30pm
Cancer biology monthly seminar	Thursday	12pm
Human genetics monthly journal club	Friday	3:30pm

Table 4. Seminar schedule during academic year.

Professor Kendziorski is teaching a special topics course on personalized genomic medicine. Boehm will be a full participant in that course. He is excited by the opportunity not only to strengthen his formal background in this new research field, but also to reach out to other researchers and classmates with overlapping interests.

Since Boehm studied at UW-Madison for both his B.S. and M.D. degrees, he has many connections to local researchers, including those in radiology, chemistry, biochemistry, oncology, internal medicine, and other departments. We fully expect him to take advantage of these connections to develop formal research collaborations as he develops into an independent scientist.

II.C.1.iv. Consulting, communicating statistics, and teaching. For Boehm, communication of statistical concepts and methods to non-statisticians is a professional and ethical obligation. We encourage his interest by providing opportunities for consulting and teaching, as such experiences will enable him to develop a style for and comfort with communicating quantitative ideas to other scholars. Towards this end, Boehm will enroll in a course in teaching statistics in the spring 2013 term. Professor Rick Nordheim, a highly regarded statistician and teacher, will lead the 15-week examination of methods for teaching statistics to diverse audiences.

Boehm enjoyed his previous experiences with teaching statistics, first as a biostatistics teaching assistant at the University of Washington and more recently as a short course instructor in statistical computing here in Madison. He intends to teach future short courses in statistical computing and looks forward to more extensive teaching experiences as a TA or co-instructor in the near future.

Additionally, Boehm, in keeping with his interest in disseminating quantitative methods to diverse populations, serves as assistant coach for a team of middle school mathematics students who will compete in the nation-wide MathCounts competition this spring. He develops training plans with which he instructs students during their weekly 2.5-hour training sessions. He plans to continue with MathCounts coaching throughout his time in Madison.

II.C.1.v. Developing skills to direct an independent research program. Boehm has developed a keen interest in learning skills that will aid in his transition to being an independent scientist. In addition to courses to strengthen his statistical skills, Boehm will enroll in multiple courses that address professional and scholarly development. Botany 575 “Scientific writing”, in which Boehm will enroll in Spring 2013, aids students in developing skills for grant writing and manuscript writing. He will extend his training in scientific writing after completion of Botany 575 through monthly progress reports, scientific manuscript preparation and submission, and grant proposal writing. Boehm intends to prepare and submit for publication at least three manuscripts during his fellowship period. These are detailed in the timeline section of the research strategy document.

To improve his public speaking skills, Boehm will continue to present annually at the toxicology seminar and twice per year at the statistics department's student seminar for all years of his training. To enhance his poster presentation skills, he will present at a minimum of two poster sessions per fellowship year. Among the venues for his poster presentations are 1) Molecular toxicology retreat and 2) Computation in biology and medicine program retreat.

Boehm wants to gain an ability to manage a team of researchers. His current and former advisors Bruce Weir, Michael Newton, Michael Gould, and Christina Kendziorski have demonstrated aptitude for this skill set, and Boehm plans to discuss research group management with them and other faculty members. Furthermore, Boehm wants to develop as a research mentor. He has many role models, including his sponsor Michael Newton and co-sponsor Michael Gould. Boehm will work with Gould and Newton to arrange for assignment of an undergraduate mentee for 10 weeks during the summer of 2014. During the summer of 2014, Boehm will enroll in the UW-Madison's 8-week course on “Becoming a research mentor” to complement his experience with an undergraduate mentee.

Finally, Boehm needs to connect with the greater biostatistical community. Towards this goal, Boehm will attend at least one scholarly research meeting or workshop per year. In August 2013, he will attend the Joint Statistical Meetings. He will attend sessions on high-dimensional statistics, Bayesian statistics, and statistical genomics to complement his research training and formal coursework. Boehm will update his meeting and workshop schedule as future meetings are announced.

II.C.2. Environment and research facilities. The University of Wisconsin is one of the nation’s leading and largest research institutions providing a rich research environment. Within the UW School

of Medicine and Public Health, the Department of Biostatistics and Medical Informatics provides a vital environment for research, collaboration, and training in biostatistics and informatics. Strong faculty-level connections support a high level of interdisciplinary research and enable the coordination of graduate student training. The Department of Statistics, founded in 1960 by Professor George Box and residing in the College of Letters and Sciences, now has 27 faculty members, has in training approximately 100 graduate students, and awards 8-10 PhD degrees annually. Graduates pursue careers in academia, industry, or government. Roughly one third to one half of the PhDs work in biological or biomedical fields. The quality and breadth of both research and teaching of the faculty are fundamental strengths of the Department. Ten of the current faculty members have attained the honor of election as Fellow of the American Statistical Association and eight that of Fellow of the Institute of Mathematical Statistics.

II.C.3. Skills and techniques to be learned

II.C.3.i. Multivariate statistics: theory & application. Multivariate statistical methodology is central to inference from high-dimensional genomic data. Boehm's training in multivariate statistics has already begun. The fellowship gives him an opportunity both to formalize his education in multivariate statistics and to deepen his understanding of multivariate methods. Much of his learning will be through independent reading under the guidance of Michael Newton. Boehm will center his studies around three texts, all of which he owns: “Modern multivariate statistical techniques” by Izenman, “An introduction to multivariate statistical analysis” by Anderson, and “The elements of statistical learning” by Hastie, Tibshirani, and Friedman.

II.C.3.ii. False discovery rate methodology and mixture models. Boehm's proposed research requires a detailed understanding of mixture models, false discovery rate methods, and related empirical Bayes methods. He will acquire this understanding through formal coursework and independent study under the supervision of Michael Newton. The formal coursework component of this task requires Boehm's enrollment in Chunming Zhang's course, “Statistics 992: Large-scale inference” in the spring 2013 term. The course will use Efron's text “Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction”. Areas of particular focus for statistics 992 include large-scale hypothesis testing, significance testing algorithms, false discovery rate methodology, local FDR, and prediction. The final project for the course involves preparation of a 10-page research proposal. Boehm will use this opportunity not only to apply and to synthesize his understanding of course material, but also to enhance his scientific writing skill set. Boehm will supplement and extend this course material with independent readings under Michael Newton's guidance beginning in September 2013 and extending until the end of his fellowship period in January 2015.

II.C.3.iii. Breast cancer molecular biology & genomics. Boehm's research will enable and will require him to develop expertise in breast cancer biology. Through ongoing review of scientific and medical literature, UW-Madison cancer biology seminar attendance (http://www.humonc.wisc.edu/index.php/Cancer_Biology_Seminar_Series), and interactions with biology trainees, such as those in the Gould laboratory and those in the molecular toxicology program, Boehm will maintain his grasp on cutting-edge science in cancer biology and genomics.

II.C.4. Relationship to applicant's career goals. Boehm's proposed training blazes a trail for his development as an independent biomedical scientist. He will study under internationally recognized experts in his discipline of choice, statistical genomics. Given his prior medical training, completing a Ph.D. in statistics will enable Boehm to work at the interface of medicine and statistics as he contributes to developments in personalized genomic medicine.

II.D.Number of Fellows/Trainees to be Supervised During the Fellowship. In addition to Boehm, sponsor Newton has 4 predoctoral students in statistics: Q. Pei, Q. He, Z. Wang, and N. Henderson. In addition to Boehm, co-sponsor Gould supervises 5 trainees: postdoctoral trainees B. Smits, C. Wiederholt, and J. Fautch, and predoctoral trainees M. Eichelberg and A. Henning.

II.E. Applicant's qualifications and potential for a research career.

University of Wisconsin
SCHOOL OF MEDICINE AND PUBLIC HEALTH

Department of Biostatistics & Medical Informatics

December 10, 2012

NIH Fellowship Reviewers

Re: Qualifications and Research Potential of Fred Boehm, MD, UW Madison

To whom it may concern:

I am delighted to comment on Dr. Fred Boehm's fellowship application. I will briefly review and endorse Fred's research plan, and emphasize why Fred is uniquely qualified to advance research in eQTL sample size planning and thereby begin a fruitful research career at this important interface between biomedical and mathematical sciences.

This fellowship application is highly worthy of support. It is positioned at an incredibly important and fertile research interface, between complex-disease genomics and biostatistics, and it is being pursued by a talented multi-disciplined new investigator. Fred Boehm, who is a medical doctor with experience in genomic analysis, is now immersed in a highly mathematical and computational PhD program biostatistics. The typical statistician, who might have the technical skills to study methodological aspects of some data-analysis technique, would not be so well-versed in critical aspects of the biomedical context driving the research forward; at the same time, the typical biomedical scientist would not have the mathematical and computational skills to bring the necessary calculations to fruition. In my 22 years of biostatistics training I have not seen any student able to align these key skills, and so I am particularly excited about Fred's efforts.

Fred's research proposal has emerged from his work over the last year with me and Dr. Michael Gould on our computationally-focused R01 to learn more about the biological function of breast-cancer-associated single-nucleotide polymorphisms (SNPs). The work relies on genotype and expression phenotype data on a valuable collection of normal breast epithelial cells. If the role of these GWAS-identified SNP's could be studied in normal tissue there could be important new knowledge gained on their cancer-risk effects. Fred's job has been to integrate the different high-dimensional data sources and apply suitable statistical processing methods. We have found some exciting preliminary signals but are continuously aware of the low signal to noise ratios. A central problem facing investigators is having a realistic system for planning sample sizes in this and similar genomic studies. I have been impressed with Fred's comprehensive analysis of the breast-cancer problem and his initial plans for study designs. He is fully capable of carrying out the proposed research and of developing a research program in this fruitful area.

Sincerely,

Box 4675, K6/446 Clinical Science Center
600 Highland Avenue
Madison, WI 53792–4675

www.biostat.wisc.edu

608/263–1706
FAX 608/265–5579

Michael A. Newton, Ph.D.
Professor, Departments of Statistics and of
Biostatistics and Medical Informatics
Co-director, Cancer Genetics Program, UWCCC