CORE B

Title: Genomics and Bioinformatics Core

W. Kelley Thomas

Principle Investigator

**SPECIFIC AIMS (GENOMICS AND BIOINFORMATICS CORE)**

Infrastructure is critical to efficient and rapid advancement in biomedical research. To be competitive, research programs require cost-effective, state-of-the-art infrastructure operated under an effective management plan. In our COBRE, access to research infrastructure is all the more critical, given the fact that the target cohort of the Genome-Enabled Biomedical Research Institute (GEBRI) are often not molecular biologists nor are they necessarily familiar with diverse bioinformatics analysis approaches, and therefore require both intensive training and support. The operation of an effective genomics core in the context of the GEBRI is not simply a matter of providing access to the instrumentation but must be accompanied by extensive training and support for experimental design and the application of molecular technologies. Similarly, while computational methods of data analysis have matured and become more user friendly, intensive support and training is required to gain the expertise needed to understand and implement those tools. To address these fundamental needs for the faculty of the GEBRI and regional partners, this proposal focuses on the infrastructure, expertise, and level of support necessary for developing the multidisciplinary research programs envisioned by this COBRE.

The long-range goal of the GEBRI Genomics and Bioinformatics Core is to establish a sustainable research support unit that provides the appropriate platforms for analysis as well as the intensive support and training necessary to foster the development of multidisciplinary research programs. This will require proactive management to ensure that the GEBRI faculty are equipped to conduct their research using best practices either through services provided by the Genomics and Bioinformatics Core at the University of New Hampshire (UNH) or through regional academic partners.

The specific aims of the Genomics and Bioinformatics core are:

**(1)** **To establish a shared genomics core facility that specializes in intensive support of multidisciplinary research groups.** This goal requires that the shared genomics facility be staffed with sufficient support personnel to shepherd users through new experiences. At the same time, the core must implement state-of-the-art technologies and provide best practices in a field that is still advancing rapidly. This will be accomplished by coupling activities outlined in the Administrative Core (Core A) to maintain an active knowledge of the field, targeted implementation of automation and single-cell genomics technologies, and by leveraging regional partners to bring a greater diversity of platforms and expertise to multidisciplinary research programs.

**(2) To** **deliver the computational infrastructure and training to ensure GEBRI faculty and regional partners are enabled to conduct bioinformatic analyses and employ the best experimental design and computational approaches**. Accomplishing this goal will leverage existing shared infrastructure and the complementary activities of a partner COBRE at Dartmouth (under the direction of Steering committee member, Jason Moore). This aim specifically intends to (1) Focus on the development of the UNH node of the NH Grid with specific capacities best suited to the needs of the proposed GEBRI programs and (2) Utilize the expertise of the Dartmouth Computational Biology COBRE (iQBS) to enhance collaborative training opportunities of the GEBRI faculty in the areas of statistical analysis, data mining, and exploratory visualization.

Having the capacity to support GEBRI faculty and regional partners from study design through data generation and computational analysis will provide the critical support infrastructure and training necessary to (1) adopt genomics approaches and develop genomic infrastructures in novel systems and (2) bring the tools of genomics and bioinformatics to research groups from disparate disciplines. These expected outcomes are particularly critical for the development of multidisciplinary research programs as envisioned GEBRI faculty.

**SIGNIFICANCE (GENOMICS AND BIOINFORMATICS CORE)**

The proposed GEBRI Genomics and Bioinformatics Core will impact many research programs in the region. The neighboring IDeA state universities and research centers house many investigators in disciplines that can have a significant impact on biomedical research when enabled with the tools of genomics and bioinformatics. For example, studies of the human microbiome arose from microbial ecology enabled with genomics and bioinformatics. The development of that multidisciplinary field is now one of the most rapidly growing and impactful areas of biomedical research. Similarly, because of its laboratory and genetic prowess the model organism *Saccharomyces* (Yeast) is known to be the single greatest source of functional annotation for the human genome. However, the tools of genomics and bioinformatics can engage many more specialized systems with unique properties and insights into biomedical investigation and bring those novel systems into the pipeline for understanding functional mechanisms important to human disease. The infrastructure and dedicated support proposed in the Genomics and Bioinformatics Core is explicitly targeted to facilitate similar multidisciplinary transformations of existing research programs and nascent model systems and has the potential to significantly impact the number of NIH-funded research programs at the University of New Hampshire (UNH) and regionally. A second significant impact of this proposal will stem from the implementation of automation for cost-effective high throughput sequencing and the incorporation of capacities for implementing single-cell genomics as new regional technologies. The implementation of these new technologies will have a significant impact on all regional genomics programs.

**INNOVATION (GENOMICS AND BIOINFORMATICS CORE)**

The proposed Genomics and Bioinformatics Core is innovative for three reasons. First, the intention of the GEBRI is to provide intensive support to research groups in disciplines other than molecular biology (e.g. chemical engineering) to foster the development of genomics and bioinformatic approaches creating explicitly multidisciplinary research programs. Multidisciplinary knowledge and skills are critical to addressing long standing questions in biomedical research and the development of new tools for treatment. However, achieving the benefits of multidisciplinary research requires the support of infrastructure and personnel dedicated to the mission of bringing these technologies into research programs expanding into genomics and bioinformatics. As described here, the Genomics and Bioinformatics Core is innovative in its levels of service and support provided to those target user groups. Second, the proposed core will implement new technologies for the implementation of genomic analysis of small samples, dramatically improving the research capacities of virtually all users of this core. Third, this Core is developed specifically to complement activities and infrastructure in the regional partners. This includes the coordinated efforts and complementary genomic platforms across institutions and the regional network of computational infrastructure (NH Grid) upon which to build. These opportunities will allow more cost-effective data gathering on a diverse set of platforms for both genomic and computational analysis.

**APPROACH (GENOMICS AND BIOINFORMATICS CORE)**

**AIM #1 To establish a shared genomics core facility that specializes in intensive support of multidisciplinary research groups.** This goal requires that the shared genomics facility be staffed with sufficient support personnel to shepherd users through new experiences. At the same time, the core must implement state-of-the-art technologies and provide best practices in a field that is still advancing rapidly. This will be accomplished by coupling activities outlined in the Administrative Core (Core A) to maintain an active knowledge of the field, targeted implementation of automation and single-cell genomics technologies, and by leveraging regional partners to bring a greater diversity of platforms and expertise to multidisciplinary research programs.

***Introduction.*** The Genomics and Bioinformatics Core of the GEBRI is designed to bring the highest quality and most cost-effective research infrastructure in genomics to the member faculty. Changes in DNA sequencing have driven the costs down, making access to genome/transcriptome data achievable by a greater diversity of research programs. Consequently, genomics is no longer the purview of large genome centers and projects in genomics and bioinformatics can be most efficiently supported by local shared facilities, if appropriately managed. As a result of the availability of sequencing platforms and the dramatically reduced costs, it is now possible to develop high quality, genomic resources for virtually any organism. However, while this activity is becoming very common, the utility of those resources is often limited by the support for both the data generation and the necessary computational support to make these genomic resources useful. Consistent with the GEBRI theme and because it specializes in the support of multidisciplinary programs and intends to engage faculty not formally trained in genomics and bioinformatics, this core must adopt a concentrated approach with a structure that allows and promotes extensive interactions and training in all aspects, including experimental design, technology, data analysis, interpretation and publication (e.g. data sharing). The aim is challenging in the face of rapidly changing technologies and the core must be prepared to adopt best practices, which maintain cutting-edge infrastructure and applications. Critical to these goals will be envisioning this core as a partner in regional Genomics Cores increasing the diversity of platforms and expertise available to the GEBRI faculty.

***Justification and Feasibility.*** The existing genomics core facility at the UNH Hubbard Center for Genome Studies (HCGS) has a history of supporting the developing genomic resources for novel systems and engaging new faculty in genomics and bioinformatics. These have included the development of the NIH model microcrustacean (*Daphnia*) as an important system for understanding the mechanisms of reaction to environmental stressors (Colbourne et al. 2011), the development of the Little Skate genome project with regional partners (Refs) and the establishment of new genomic infrastructure for the study of circadian rhythms in novel species undertaken by primarily undergraduate institutions (PUI) faculty (Chabot, Newcomb and Watson) supported in the New Hampshire INBRE program. However, the capacity to support additional training as described for this center is challenged by the intensity of the interactions necessary to support this user group which require significantly more time than simply running a shared core for data generation. Our target cohort of faculty brings significant disciplinary expertise. When that is combined with genomics and bioinformatics it unlocks new opportunities to address fundamental, biomedically important questions. However, the key is to enable faculty to adopt genomics and bioinformatics and achieve a multidisciplinary approach to their scientific question.

A second major challenge to providing support for Genomics is that the technology continues to change at a rapid pace. The HCGS currently operates traditional sequencing and genotyping (ABI capillary sequencing) and high throughput sequencing on the Illumina HiSeq 2500 platform, with all traditional peripheral library generation technologies in place. Under the direction of PI Thomas the HCGS has maintained an active genomics core facility since opening in 2001. Over that time, the HCGS maintained access for users to state-of-the-art instrumentation through the major changes in DNA sequencing. This included early adoption of Next Generation sequencing platform (Refs—from Kelley’s cv). While the HCGS currently has excellent sequencing instrumentation, advances in technology (specifically microfluidic library preparation) are dramatically reducing sample preparation efforts and costs while at the same time increasing consistency in quality across samples and reducing the necessary amount of input material further expanding the types of samples that can be subjected to sequencing based assays.

In addition to the automation described above, among the most important recent advances in genomics are the new developments in the use of extremely small sample quantities, so called single-cell genomics (Blainey, 2013; Shapiro et al, 2013). The ability to sequence genomic DNA or RNA from a small number of cells can fundamentally change the approaches taken to assay both genotypes and patterns of expression. By reducing the course population-based approach to genomics looking at signals shared across the entire population, the ability to reduce that population size to a single cell eliminates the loss of signal in the noise of neighboring cells, allowing true cell linage relationships to be derived and supplants the use of marker-based definitions of cell types. Finally, maintaining both state-of-the-art and appropriate platforms for diverse applications is impractical within a single core and requires working closely with the partner institutions to provide infrastructure and expertise.

The existing genomics and bioinformatics core facility at UNH, the HCGS, is directed by PI Thomas who has over 20 years of experience directing shared DNA sequencing facilities. PI Thomas has also been instrumental is the theme-specific application of new sequencing technologies to his ongoing NIH-funded research programs focused on elucidating rates and patterns of mutation and more recently, has developed novel approaches for the analysis of error rates in transcription (Gout et al. 2013). The HCGS facility is managed by Krystalynne Morris (BA, MBA) who oversees all operations. Ms. Morris has over 15 years of experience managing a shared genomics facility. Bioinformatics support is coordinated and overseen by Feseha Abebe-Akele (Ph.D.) and bioinformatics staff focused on activities of the NH INBRE bioinformatics core, supporting multiple PUIs in the state of NH.

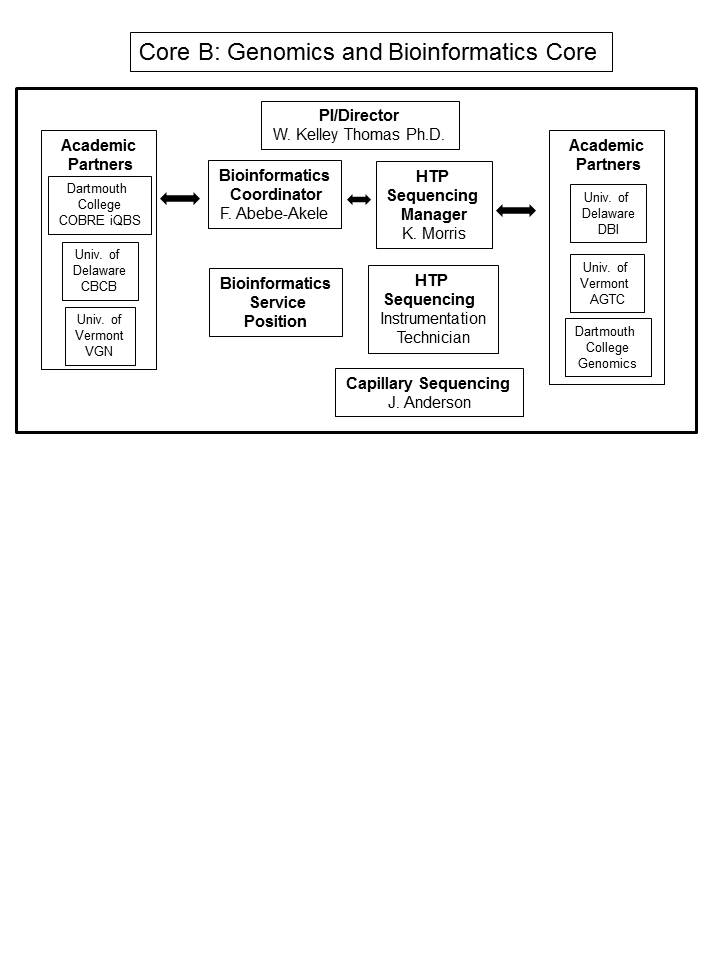


Figure 1. The organizational chart for the proposed Genomics and Bioinformatics Core. Academic partners include the Institute for Quantitative Biomedical Sciences (iQBS) at Dartmouth College, the Center for Bioinformatics & Computational Biology (CBCB) at the University of Delaware, the Vermont Genetics Network (VGN) at the University of Vermont, the Delaware Biotechnology Institute (DBI) at the University of Delaware, the Advanced Genome Technologies Core (AGTC) at the University of Vermont and the Dartmouth Genomics Shared Resource (Genomics) at Dartmouth College.

***Design.*** To accomplish the goals of AIM 1, there are two specific sub-aims.

**AIM #1a** **To expand the capacity of the genomics and bioinformatics core to support the GEBRI faculty.** To facilitate intensive support, the capacity of the genomics core will be expanded by hiring an additional full-time technician with specific duties to implement sample preparation and automation within the core. Currently, the bottleneck in the existing facility stems from researchers’ needs for extensive support in sample preparation and library construction. This level of support is critical for the generation of high quality data and the technologies involved do not need to be housed in individual research labs not generally engaged in molecular biology. The additional staff will free up time for more consultation by the sequencing core manager and training of GEBRI faculty in experimental design through data analysis and publication support necessary to achieve independent research capabilities. The sequencing core manager will be able to devote 25% FTE effort specifically to this activity.

**AIM #1b**  **To implement new technologies in the genomics and bioinformatics core.**  Ongoing changes in sequencing strategies require constant upgrading of facilities. At this time, to maintain the ability to provide meaningful, cost-effective support, the Genomics and Bioinformatics Core must incorporate both automation and the capacity to generate sequences from very small samples, including single-cell sorting into the core facility. Automation, specifically new microfluidic approaches to library construction, significantly reduce both costs and technician time. Current processes in the HCGS incorporate automation only for highly multiplexed samples. Such experimental designs are becoming more prevalent as population genomics and increased replication within experiments are becoming the normal standard, increasing the demand for highly multiplexed sequencing library preparation. For the past year, previous automation has been done in a partnership with Indiana University. However, while the HCGS continues that relationship, the automation at Indiana University does not include microfluidics, which is the key to dramatic cost reductions. To bring these technological advances and cost reductions to the GEBRI faculty, the Genomics and Bioinformatics Core will incorporate a platform for microfluidic automation with the capacity to generate 384 libraries in parallel (e.g. CLiC genomics)

The ability to generate high quality sequence libraries from small samples is possible using new automation methods and by the incorporation of cell sorting technology into the core. The use of small samples dramatically impacts GEBRI faculty because many novel systems are single-celled and many biological functions are restricted to small numbers of cells. Genomic analysis on populations of cells, as is the previous standard, inherently limits the detection of signal among the critical cell subpopulations. These improvements in infrastructure will immediately impact the GEBRI faculty and regional partners. The Genomics and Bioinformatics Core will implement this goal by establishing cell-sorting instrument that has broad capabilities (FACSJazz) to provide high quality support for the broadest array of user applications including both eukaryotic and prokaryotic systems.

While the additional technical support, automation, and support for small samples will be critical infrastructure for GEBRI faculty and regionally, it is recognized that GEBRI faculty will best be served by coordinating efforts among regional cores. For the last 5 years the Core Directors from this region have met annually in the context of the local core facility meetings. The North East Regional Life Sciences Core Directors (NERLSCD) meeting has been an outstanding opportunity for the IDeA state cores to meet and strategize the advancement of these facilities and to discuss the challenges of implementing shared research resources. Among the critical outcomes of this collaboration among facilities is the sharing of the unique sequencing platforms available among academic partners (Figure 1). In New Hampshire, there are extensive shared resources housed in the Norris Cotton Cancer Center Geisel School of Medicine at Dartmouth. These are coordinated by academic partner Stephen Bobin and include Ion Torrent Proton and Personnel Genome Machines in their Genomics core directed by Craig Tomlinson. The HCGS has a long standing collaboration with the Center for Bioinformatics and Computational Biology at the University of Delaware directed by Cathy Wu. Their genomics core facility, directed by Bruce Kingham, operates a Pacific Biosciences instrument and other sequencing platforms. In addition to the unique platforms maintained by partners, the HCGS works with a sister core facility at the University of Vermont (The Advanced Genomic Technologies Center directed by Tim Hunter and operated by Scott Tighe) to coordinate Illumina-based sequencing to minimize wait times and allow focused application support. This assemblage of sequencing systems in the region supports a diversity of unique applications (e.g. in marker-based metagenomics (ION) and large-scale genome assembly (PacBio)). In particular, GEBRI faculty developing genomic resources for novel organisms will benefit from access and expertise in the use of the PacBio system at the University of Delaware facility (with bioinformatics support by academic partner Shawn Polson) which is developing as a key player in novel genome assemblies. It is critical to emphasize that these partners represent both the instrumentation and expertise (including bioinformatics) specific to each platform.

***Expected outcomes.*** When combined with intensive support and extensive training, the implementation of automation, techniques for single cell genomics, and shared resources with partner institutions will provide for the cost-effective and high quality support of genome-enabled research programs of GEBRI faculty.

**AIM #2 To** **deliver the computational infrastructure and training to ensure GEBRI faculty and regional partners are enabled to conduct bioinformatic analyses and employ the best experimental design and computational approaches**. Accomplishing this goal will leverage existing shared infrastructure and the complementary activities of a partner COBRE at Dartmouth (under the direction of Steering committee member, Jason Moore). This aim specifically intends to (1) Focus on the development of the UNH node of the NH Grid with specific capacities best suited to the needs of the proposed GEBRI programs and (2) Utilize the expertise of the Dartmouth Computational Biology COBRE (iQBS) to enhance collaborative training opportunities of the GEBRI faculty in the areas of statistical analysis, data mining, and exploratory visualization.

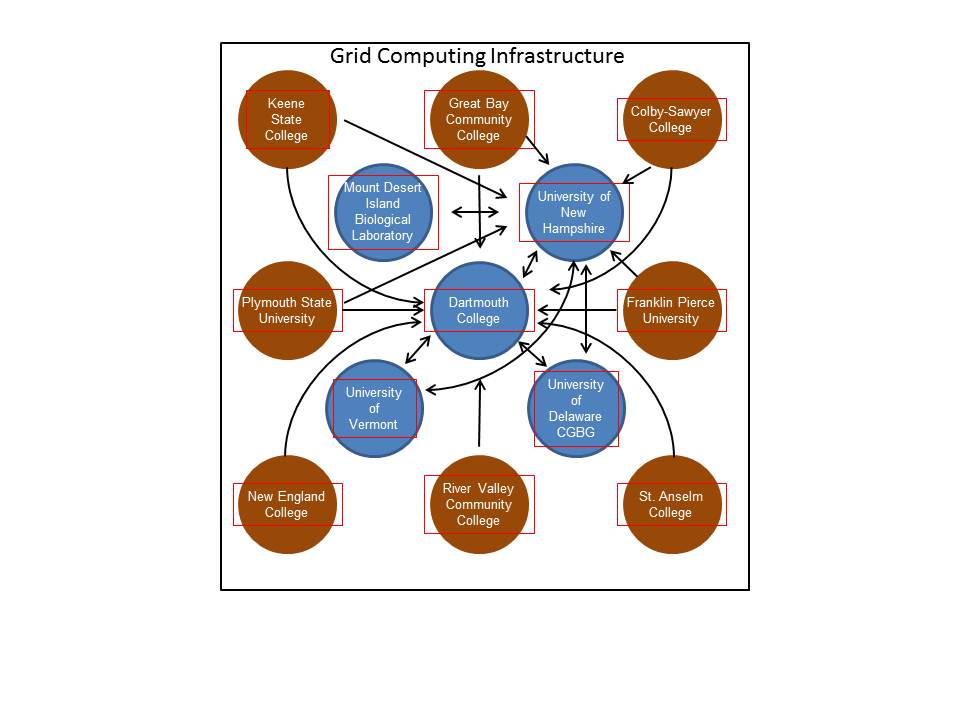
***Introduction.*** Bioinformatics and genomics are inextricably linked. The technological advances in genome sequencing would be of little value without parallel development in the methods of analysis. Fortunately, tremendous progress has been made in recent years to facilitate data analysis and data handling. Nevertheless, as data generation changes in scope and intrinsic qualities, bioinformatics approaches rapidly change to meet the new challenges. Like the molecular biology aspects of genomics, bioinformatics is a challenging yet necessary skill set to take advantage of its disciplinary values. Bioinformatics includes multiple key aspects from data processing to computational biology and understanding the implications of patterns. Bringing access to these approaches to the GEBRI faculty requires establishment of the infrastructure, training and partnerships with programs having complementary expertise in computational biology. As in Aim #1, effective bioinformatics support will enable GEBRI faculty to engage these approaches in their research groups. Achieving that level of training will require dedicated and intensive support of this core.

***Justification and Feasibility.*** The Genomics and Bioinformatics core facility in the HCGS at UNH provides support for the analysis of all genomics projects conducted through the HCGS sequencing core. Direct support is provided by a full time coordinator of Bioinformatics (Feseha Akele-Abebe, Ph.D.) and multiple, part-time assistant programmers. These staff are supported by the College of Life Sciences and Agriculture (COLSA) and the Vice Provost for Research office (Dean Wraith and VP Nisbet, respectively). In addition, one full time bioinformatics staff member is supported in the bioinformatics core facility at UNH as a component of the New Hampshire-INBRE (PI Ron Taylor, Dartmouth). Existing staff have two primary duties. First is direct support and consultation on individual projects. The bioinformatics coordinator and staff are all involved in experimental design and analytical implementation from the initial consultation. Second, each member of the bioinformatics team is expected to participate in outreach through the development of training workshops to support common applications. The proposed GEBRI Genomics and Bioinformatics Core will continue and expand those practices.

One of the key aspects of support for bioinformatics at UNH is a curriculum for training developed with Computer Science faculty academic partners Dan Bergeron (Ph.D.) and Phil Hatcher (Ph.D.). This includes a formal course in bioinformatics targeted to biologists and an applied bioinformatics course pairing life science students and computer science students to develop novel applications and pipelines. Both of these courses serve a critical research support function by training students, technicians, postdocs and faculty in core bioinformatics activities. PI Thomas, and another MCBS faculty member have recently led the creation of a new multi-disciplinary doctoral program in Molecular and Evolutionary Systems Biology whose thematic focus aligns extremely closely with the COBRE-GEBRI scientific mission. Sustainable GEBRI research programs will require access to such a trained workforce. These curricula complement opportunities developed within the NH INBRE Bioinformatics Core and with partners offering application-specific bioinformatics workshops. In the past 5 years, there have been numerous opportunities to share in both the development and delivery of workshops in bioinformatics with regional partners. Specifically, together with partners at Mount Desert Island Biological Laboratories (partner Ben King) and the Bioinformatics and computational Biology program at the University of Delaware (partner Kathy Wu and Shawn Polson), the HCGS staff participated in a series of bioinformatics workshops focused on the sequencing and annotation of the little skate genome (King, et al, 2011; Wang et al. 2012), an exemplar of our theme. The HCGS continues to participate in the annual bioinformatics workshops in environmental genomics offered by MDIBL. Similarly, with the Bioinformatics Core at the University of Vermont (with partner James Vincent, director of the Vermont Bioinformatics core), the HCGS has offered workshops in metagenomics. As part of its role in the NH INBRE, staff of the HCGS routinely deliver workshops on genomics and bioinformatics to regional PUI laboratories.

Existing infrastructure for bioinformatics includes a large scale (25 TB) secure data storage facility operated as part of our research computing facility (Director, Patrick Messer). This facility also hosts the UNH node of the New Hampshire Grid, a network of high-performance computers established to provide shared computing infrastructure to NH and regional IDeA state and EPSCoR state partners (Figure 2).

These same resources serve NH INBRE partners and those in neighboring states. For both storage and computational needs, shared resources provide a cost-effective opportunity to have the necessary computational power when needed at a fraction of the cost. Like the genomics technology partners, the regional bioinformatics cores provide opportunities to share both infrastructure and a greater diversity of expertise.

Figure 2. Components and connections for the participating institutions in the NH Grid. Blue circles indicate participating research cores and the brown circles represent PUI institutions in the NH-INBRE.

Together this infrastructure and expertise comprise a valuable resource for our GEBRI faculty. However, the specific applications that are inherent in the proposed work involve a variety of optimal computational configurations. The existing network allows for massively parallel applications but includes only a minority of nodes with the high memory requirements needed for a number of specific applications, such as genome assembly, which result in delayed computational processes.

Improvements in cyberconnectivity make this computing grid even more valuable. Since 2007, significant investment in the regional cyberinfrastructure has dramatically improved connectivity among regional partners and the rest of the world. Most recently, improvements on the UNH campus (NH Broadband Action Plan, NIH NCRR COBRE Supplement (Dartmouth College) and NSF EPSCoR RII Track 2 (UNH Internal Advisory Committee member Scott Valcourt)) are creating the final critical aspects of connectivity that will allow for more seamless sharing and distributed computation as envisioned for the NH Grid.

***Design.***

In order to establish a highly effective computational infrastructure, the GEBRI Genomics and Bioinformatics Core will specifically build on the existing shared resource (the NH Grid). This approach connects the proposed expanded resource with a large-scale distributed computing cluster, including high performance platforms throughout the state. The cluster began with an initial investment by steering committee member Jason Moore (Dartmouth College) and is currently support by contributions from the Institute for Quantitative Biomedical Sciences COBRE at Dartmouth, the New Hampshire INBRE, as well as other partner investments. The motivation for this approach to building the necessary computational infrastructure is that by building the resource to serve GEBRI-specific applications, to the extent those needs are shared by others in the NH Grid, this infrastructure will be valuable to other regional research groups. The GEBRI-specific additions to the NH grid will focus on High RAM nodes. These will be housed at UNH in the data center and made available as part of the infrastructure to maximize opportunities for sharing.

The expanded bioinformatics needs of the GEBRI members will require additional, dedicated bioinformatics support. This need will be met by the hiring of an additional full-time bioinformatics support staff person. This new staff member will be specifically designated to support the training and implementation of bioinformatics for the GEBRI faculty. The UNH Vice Provost for Research and the Dean of COLSA will continue to support the current bioinformatics staff for the duration of the COBRE.

***Expected outcomes.*** The expected outcomes for this aim of the proposed activities in the Genomics and Bioinformatics Core include the establishment expanded computing infrastructure connected to a large-scale shared facility as well as increased bioinformatics support and training. Planned further expansion of regional opportunities for training and support in Bioinformatics will also facilitate the research of GEBRI faculty.

***Potential problems and alternative strategies.***

The two greatest potential problems faced by the proposed GEBRI Genomics and Bioinformatics Core are (1) the need to provide effective training and meaningful support to non-molecular biology programs; and (2) rapidly changing technologies. Both genomic and bioinformatic technologies change at a pace that belies accurate predictions for 5 years except that the technology will be different, cheaper and better. In order to maintain a valuable resource, the GEBRI Genomics and Bioinformatics Core and partner institutions are committed to maintaining support of the best practices of our faculty. The plans to expand personnel and a commitment to maintaining state-of-the-art infrastructure combined with cooperation among regional partners will mitigate these challenges. It is time-consuming but critical to bring these powerful technologies to bear on important biomedical questions.

Similarly, the rapidly changing scope and qualities of genomic data make long term planning for storage and platforms for processing a challenge. This proposal leverages existing infrastructure in a federated approach hedging against massive mis-investment. Nevertheless, planning data storage and processing approaches for five years is challenging and will likely involve further implementation of cloud computing strategies for analysis already common for a handful of stable applications as well as for long-term secure data storage. Together with the academic partners and under the advice of both the internal and external advisory committees, this core will continue to monitor for future best practices.

**Expected Contributions to the individual COBRE GEBRI projects**

Each of the four GEBRI projects will benefit from the establishment of the Genomics and Bioinformatics Core.

Project 1: **Kyung Jae Jeong, Ph.D.** Assistant Professor, Department of Chemical Engineering**, focuses on tissue bioengineering by manipulating stems cells and substrates to produce functioning tissues. Dr. Jeong will benefit from the GEBRI Genomics and Bioinformatics Core by having access to state-of-the-art genomic and bioinformatics resources and training with a high level of service. The use of a genomics and bioinformatics approach is critical for understanding the biological mechanisms for the stem cell response to different engineered substrates. Dr. Jeong’s research will specifically involve the use of RNA-seq to compare transcription profiles and bisulfite sequencing to examine changes in methylation patterns in the stem cell populations studied. This program is explicitly multidisciplinary and brings valuable knowledge of stem cell biology, bioengineering and significant biomedical applications to this GEBRI.**

Project 2: Kang Wu, Ph.D., Assistant Professor, Department of Chemical Engineering, focuses on the development of novel microbial drug delivery systems. Dr. Wu’s work requires extensive comparative genomics in the experimental design and proteomic and bioinformatics resources to accomplish the overall goals. Dr. Wu will specifically use high throughput DNA and RNA sequencing as well as mass spectroscopy approaches available in the GEBRI Genomics and Bioinformatics Core. Dr. Wu bring an important new dimension to our excellent research group in microbial genomics and will benefit significantly from available support for the development of her capacity to undertake the genomic and bioinformatics aspects of her proposal. Dr. Wu’s research program is explicitly multidisciplinary and brings valuable engineering and synthetic biology perspectives to this GEBRI

Project 3: Matthew MacManes, Ph.D., Assistant Professor, Department of Molecular, Cellular, & Biomedical Sciences, focuses on the elucidation of novel mechanisms of tolerance to dehydration in mammals using comparative and functional genomics and physiological assays in a new, powerful rodent model system. Dr, MacManes is enabled by the genomics technology and bioinformatics infrastructure in the proposed GEBRI core as well as facilities (PacBio) in the core of a regional partner. Dr. MacManes specifically requires high throughput sequencing for RNA-seq and bisulfite sequencing. Dr. MacManes’s research program is explicitly multidisciplinary and he brings to the GEBRI significant skills in bioinformatics.

Project 4: David Plachetzki Ph.D., Assistant Professor, Department of Molecular, Cellular, & Biomedical Sciences, investigates the genetic basis and developmental consequences of self/non-self recognition in a remarkable colonial animal that is uniquely suited for this research. The proposed study has important implications for our understanding histocompatibility, adaptive immunity, and stem cell biology. Dr. Plachetzki will utilize genomics and bioinformatics resources in the GEBRI core, specifically for assays of transcriptional regulation. Dr. Plachetzki’s research program is explicitly multidisciplinary and brings significant experience in comparative genomics and developmental biology to the GEBRI.

Timeline: The additional genomics technician and bioinformatics support staff will be implemented in year 1 and continue throughout the tenure of the Genomics and Bioinformatics Core. Automation and single-cell genomics platforms are established in year 1 and purchased over three years. It is anticipated that additional infrastructure will be required in years 4 and 5. NH Grid enhancement has a target of 256 cores with purchases front-loaded in with high RAM nodes in years 1 and 2.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Project Year** | **1** | **2** | **3** | **4** | **5** |
| Aim 1 Genomics |  |  |  |  |  |
| 1a Genomics staff | x | x | x | x | x |
| 1b automation and single-cell genomics | x | x | x | x | X |
| Aim 2 Bioinformatics |  |  |  |  |  |
| Staff support | x | x | x | x | x |
| NH Grid enhancement | xxx | xx |  |  |  |

Future directions

References

Blainey. P.C. 2013. The future is now: single cell genomics of bacteria and archaea. ***FEMS Microbiol. Rev.*** 37:407-427. doi: 10.1111/1574-6976.12015. Epub 2013 Feb 11. PMCID: PMC3878092

Colbourne, J.K., M.E. Pfrender, D. Gilbert, W.K. Thomas, A. Tucker, T.H. Oakley, S. Tokishita, A. Aerts, G.J. Arnold, M.K. Basu, D.J. Bauer, C.E. Caceres, L. Carmel, C. Casola, J.H. Choi, J. Detter, Q. Dong, S. Dusheyko, B.D. Eads, T. Frohlich, K.A. Geiler-Samerotte, D. Gerlach, P. Hatcher, S. Jogdeo, J. Krijgsveld, E.V. Kriventseva, D. Kultz, C. Laforsch, E. Lindquist, J. Lopez, J.R. Manak, J. Muller, J. Pangilinan, R.P. Patwardhan, S. Pitluck, E.J. Pritham, A. Rechtsteiner, M. Rho, I.B. Rogozin, O. Sakarya, A. Salamov A, S. Schaack, H. Shapiro, Y. Shiga, C. Skalitzky, Z. Smith, A. Souvorov, W. Sung, Z. Tang, D. Tsuchiya, H. Tu, H. Vos, M. Wang, Y. Wolf, H. Yamagata, T. Yamada, Y. Ye, J. R. Shaw, J. Andrews, T.J. Crease, H. Tang, S.M. Lucas, H.M. Robertson, P. Bork, E.V. Koonin, E.M. Zdobnov, I.V. Grigoriev, M. Lynch, J.L. Boore. 2011. The Ecoresponsive genome of *Daphnia pulex*. ***Science*** 331:555-561.PMCID: PMC3529199.

Gout, J.-F., W.K. Thomas, Z. Smith, K. Okamoto, M. Lynch. 2013. Large-scale detection of in vivo transcription erros. ***Proc. Natl. Acad. Sci.*** early release doi:10.1073/pnas.1309843110**.** PMCID: pending

King B.L., J.A. Gillis, H.R. Carlisle, R.D. Dahn. A natural deletion of the HoxC cluster in elasmobranch fishes. ***Science.*** 2011 Dec 16;334(6062):1517. PubMed PMID:[22174244](http://www.ncbi.nlm.nih.gov/pubmed/22174244)

Wang, Q.\*, C.N. Arighi\*, B.L. King\*, S.W. Polson\*, J. Vincent\*, C. Chen, H. Huang, B. Kingham, S.T. Page, M.F. Rendino, W.K. Thomas, D.W. Udwary, C.H. Wu, and the North East Bioinformatics Collaborative Curation Team. 2012. Community annotation and bioinformatics workforce development in concert – Little skate genome annotation workshops and jamborees. ***Database***. doi:10.1093/database/bar064. \* authors contributed equally. PubMed PMCID:[PMC3308154](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3308154)

Shapiro E, T. Biezuner, S. Linnarsson. 2013. Single-cell sequencing-based technologies will revolutionize whole-organism science. ***Nat. Rev. Genetics*** 14:618-30. doi: 10.1038/nrg3542. PMID: 23897237