

Genome Biology: Understanding the Responsive Genome in Health, Agriculture, and in the Environment (October 25, 2018 DRAFT)

Executive Summary

Genome sequencing has been practiced for less than two decades, and the impact throughout the life sciences has been HUGE. Our resulting understanding of genomes has produced explosive growth in our discovery of **how organisms develop, function, resist or succumb to disease, and age, as well as how they interact with each other and with the environment.** Genome biology is providing new tools and approaches that are transforming our fundamental understanding of how genomes are regulated, replicated and maintained. Moreover, genome biology provides the novel experimental and computational tools needed to better understand and prevent disease, and to more efficiently harness useful traits in plants and animals. Here we describe the central importance of genome biology to our research in basic science, medicine, agriculture, and ecology at Cornell, as well as our strengths and weaknesses in Genome Biology, recommended areas of future growth, and specific steps that need implementation. Also, we note that genome biology interfaces with a broad spectrum of basic studies in life sciences, and it can be further enhanced by an active interplay with physical sciences. Genome biology will synergize with ongoing and emerging Cornell Initiatives. Our specific recommendations include:

- 1) Shore-up existing areas of strength in gene regulation, DNA replication and repair, evolutionary genomics, and computational genomics.
- 2) Make judicious hires in key emerging areas of genome biology using university-wide rapid recruitment strategies that tap a pool of start-up funds and enlists more active participation of the Development Office.
- 3) Maintain a highly-efficient and adaptive infrastructure for data collection (e.g. DNA sequencing) and analysis, and create centralized training in genomics and in the dispersal of cutting-edge genomic technologies across campuses.
- 4) Create a seed grant system to fund development and use of cutting edge genomic technologies that interfaces tightly with recommendation 3 for dispersal of technologies.
- 5) Generate administrative support for center grants to develop genome technologies.
- 6) Bolster ongoing campus efforts (e.g. cancer, reproductive biology), as well as develop new opportunities, for establishing collaborative links to Weill Medical School in areas of well-funded and impactful science at the disciplinary interfaces of our campuses.*
- 7) Provide more competitive support for recruiting outstanding graduate and post-graduate researchers to the life sciences and the genome biology initiative.*

* These are being covered by broader University programs. but we emphasize the need is especially important for the Genome Biology Initiative.

The Central Importance of Genome Biology

Genome biology is both a collection of tools and a new way of thinking about biological processes and their regulation. The tools possess amazing capacities to discern changes in gene structure and function across the entire genome. The questions that these tools open up are similarly grand -- providing in one experiment, for example, quantitative measures of changes in expression of every gene in the genome through the course of development. Convergence of genome sequence data, functional genomic data, new technologies, and computational tools provide unprecedented ability to dissect biological processes. Thus, Genome Biology has brought a revolution across many of the life science disciplines, and provides an opportunity for rapid progress in some key areas outlined below.

- 1. Understanding basic biological processes.** The tools of genome biology have produced a massive acceleration in the rate of discovery of how organisms use and regulate expression of their genomes in response to developmental programs, nutritional and hormonal cues, and environmental stress. These tools and approaches are also leading to a broader understanding of mechanisms of genome replication, recombination, and repair. In addition, studies of the genomic basis for evolutionary differences across species has led to new insights into the importance of biological diversity for adaptation and speciation.
- 2. Human health and reproductive health.** Advances in genome biology have played a critical role in understanding how differences in gene sequences mediate variation in risk of complex disease in humans. These genome-wide association studies (GWAS) have been augmented by efforts to map variants that influence the expression level of genes, the methylation state of the genome, and origins of DNA replication. RNA-sequencing reveals quantitative measures of transcript abundance of genes, and has helped to develop many novel insights about tissue-specific functions in healthy and diseased states, including maintenance of stem cells. Cornell has played an important role in developing methods for human disease gene mapping as well as in understanding basic principles of genetic variation, including SNP ascertainment bias, impact of past admixture and demographic changes, epigenetic diseases, regulation and interaction of genes and environment, and epistasis. Personalized medicine is an attempt to make use of genome sequence and other “omics” data for optimal health, and the generous NIH grant centered at Weill to develop a large sample base for this work motivates many investigators across Cornell to consider how they can contribute to this effort.
- 3. Plant and animal breeding and sustainable agriculture.** Advances in plant and animal breeding in the 20th century did much to meet the food, fiber, feed, and fuel needs of the burgeoning world population. As population growth continues, the amount of land available for production decreases, and extreme weather events such as drought and flooding become more prevalent in this century, continuing to meet these needs will be an ever more challenging task. In addition to enhancing productivity and quality in an

agricultural context, adapting plants to increase their capacity for environmental remediation and improving animals as models for human disease will also help solve pressing problems. The needed advances will be driven by genome biology to link genotype-environment interactions to phenotype, understand genetic and epigenetic regulation of gene expression, and gain ever more precise control over genetic content.

4. **Ecology, global change and the environment.** Genome biology is providing tools for rapid assessment of the diversity of microbial ecosystems and for assessing the patterns and rates of change in those communities. It provides a means for examining the consequences to population decline in a wide range of organisms, and may suggest approaches to ameliorate loss of genetic diversity through breeding programs. Shifts in patterns of genetic variation have provided clear documentation of the biotic impact of global climate change.
5. **Neurobiology, behavior and the brain.** Genetics has always been a powerful tool in understanding neurogenesis, how the brain functions, and how failure of key neuronal processes result in a variety of devastating impairments to nervous system development as well as neurodegenerative diseases. Genomics ramps up the power of these approaches and often reduces the leap from model organisms to human applications.
6. **The engineering-chemistry-biology interface.** Innovative applications of massively parallel DNA sequencing have provided efficient ways to define precisely the primary sequence of genomes, the complete transcriptome, the location of specific protein-DNA interaction genome-wide, and the 3D architecture of the genome. These powerful methods and entirely novel approaches can be greatly enhanced by tapping the expertise at Cornell in chemistry, physics, and engineering. The depth, quality, and collaborative nature of these fields at Cornell provide opportunities to drive the field of genome biology to new levels. For example, new chemical methods are being developed for examining specific interactions of proteins with RNA and DNA sequences and the precise distance relationships of these components, information that is critical in understanding mechanisms that govern the regulation of genes and the replication and repair of the genome. Likewise, novel physics and engineering approaches can provide the means of observing individual cells and their differences at high efficiency and sensitivity both biochemically and optically - a view that is critical to understanding cellular interplay in basic genome biology and medicine.

Key Strengths in Genome Biology at Cornell

Cornell has an academic culture that has always been conducive to interdisciplinary and collaborative studies. With our 23 academic departments in the life sciences across five colleges in Ithaca, we have depth in a wide array of disciplines that encompass many facets of genome biology. These life science departments work in close collaboration with chemistry, physics, computer science, biomedical and other engineering departments. Notably, Cornell is exceptional among its peer Universities in that it is home to two distinct colleges that focus on

health and disease, the Medical College and the College of Veterinary Medicine. It is unique also as the sole Ivy League university that is also a land grant institution globally recognized for its excellence in plant and animal science. Additionally, the graduate field system at Cornell encourages participation of faculty from different departments in the education of graduate students. This collaborative approach can also be tapped to foster interaction across campuses, to tap into complementary strengths. Thus, the Genome Biology initiative will be designed to take advantage of our interdisciplinary opportunities and our strong desire to develop new tools and approaches to study key biological questions.

The Cornell Genome Biology Initiative can best serve the University mission by recognizing the key strengths that are already established at Cornell, and through this assessment identify key gaps and areas that are resource limited. Among the many strengths at Cornell, the four areas listed below are particularly relevant to genome biology. These disciplines present new opportunities for augmentation of strength as described in a subsequent section of this document on “Areas for Future Growth.”

1. **Genomics of gene expression:** The control of gene expression is critical for development, response to environment, and homeostasis in all organisms. The development of an organism from a single cell to an adult requires exquisite regulation of complex programs of gene expression, where key regulatory proteins interact with the genomic elements to elaborate cascades of regulatory events that in turn produce a corresponding series of highly-integrated steps that give rise to cell differentiation. Additionally, cells at all stages of development need to be responsive to nutritional, hormonal, and environmental changes. Over the past half century, focused studies of individual genes have provided considerable understanding of the molecular players (proteins, RNA and DNA elements, and epigenetic modifications) and how they collaborate to regulate gene expression. With the development of high-throughput sequencing technologies that reveal the complete sequence of genomes rapidly and at low cost, and the emerging technologies to examine expression and molecular interactions across genomes at single-base resolution, we are poised to understand in molecular detail how the genome orchestrates highly-integrated patterns of gene expression through development and in response to the environment. Because many animal and plant diseases result from mis-regulation or damage of the genome, a comprehensive understanding of genome structure and function at molecular resolution is critical. Uncovering the molecular basis of processes in normal and disease states provides the necessary insights for developing precisely-targeted therapies for diseases.
2. **DNA replication and repair:** High fidelity replication of the genome is critically important for proper development, normal cellular function and homeostasis, longevity, and reproduction. Faithful repair is also essential, as DNA is highly susceptible to damage resulting from both metabolic and environmental factors, with as many as 1 million mutagenic events per cell per day. New genomic technologies are revealing integrated views of how replication and repair of genomes takes place, consequences of failures in those processes, and ways in which DNA repair pathways can be harnessed

and manipulated to engineer genetic content.

3. **Evolutionary genomics:** Understanding evolution itself is a critical aspect of genome biology. What forces modulate levels of variation? How has past action of natural selection shaped genomes? Such inquiry informs our fundamental understanding of the stability and transmission of genetic information, gene function in the context of adaptation, and the diversity of life. Advanced computational modeling and simulation approaches are being applied to quantify and understand the processes of mutation, drift and natural selection through contrasts of multiple whole-genome sequences of related organisms. These results often dovetail with our understanding of functional constraints on components of the genome, including those that impact human disease.
4. **Computational genomics:** Big data sets, comprising complete genome sequences, functional annotation, expression data, data about variation across populations and over time, and other types of information, are fostering rapid advances in basic and applied sciences, including human health, plant and animal breeding, and environmental biology. Such data sets are made useful only through development and application of computational methods that integrate and extract the information in well defined ways. The development of genome-scale data sets must be matched by innovative advances in computation to ensure our ability to leverage genomic data for basic and applied science.

Key Weaknesses in Genome Biology at Cornell

Despite Cornell's many strengths in genome biology, significant weaknesses are limiting the level of success. Judicious investment in a few key resource-limited areas will leverage the efforts of the many disparate units where genomics is practiced. Here are some of these opportunities.

1. **Threats to existing areas of strength.** Strong disciplines like Gene Regulation, DNA Replication and Repair, and Evolutionary Genomics do not stay strong without continuing attention to maintenance of that excellence. This means not only providing competitive retention packages but also new hires in these areas. They will not stay strong without this vital renewal. One particular area that is a strength but is seriously undersized for a university of this caliber is Computational Genomics. We have produced a series of stars in this area, only to have them hired away. Structural changes in the organization may be needed to accommodate this expansion. Another area that could be considered under-supported at Cornell is stem cell biology. The potential termination of NYSTEM support would leave our fledgling program in jeopardy. Stem cell biology is central to understanding development and its application in biomedical sciences strongly motivates a senior hire in this area.
2. **Impediments to hiring in key emerging areas of genome biology.** Genome Biology is a fast-moving area, and vital tools and technologies need to be rapidly brought on

board. The approach to faculty hiring at Cornell is inherently conservative, demanding that departments have five-year plans and establish how each position fits in with that plan. Most of the time this works well, but it makes us far less nimble than many peer institutions that can quickly do cluster hires in hot emerging areas. We also need a mechanism to establish cross-departmental and cross-campus opportunities for such hires.

- 3. Inadequate training in genomics and in the dispersal of genomic technologies across campuses.** An amazing strength of the life sciences at Cornell is the diversity of organisms and life processes being studied. Many of these are being pursued with decades-old methods. Incorporation of genomics technologies into these studies could provide a major re-invigoration of this work. In the absence of a mechanism to bring these investigators on board with genomics, it is all too easy for them to continue working in the ways that they know, and lose out to others who have managed to learn new approaches.
- 4. Shortage of seed grants specifically directed to the development and application of new genomic technologies.** Having first access to the latest technology would provide Cornell researchers with a competitive edge for making new discoveries and securing external funding. A seed grant funding mechanism directed specifically to genome technology development could stimulate local efforts to develop these new tools. By tightly integrating the seed grant discovery of new technologies with the mechanisms of dispersal (point 3), the entire Cornell Community could rapidly enjoy the benefits of these new technologies. Additionally, the availability of seed funds to support application of these technologies to new projects will fuel basic and applied research and allow our faculty to be more competitive in their pursuit of external funding.
- 5. Lack of center grants to develop genome technologies.** Investigators in the life sciences at Cornell pride themselves in their individual resourcefulness, but often fail to see the advantages of larger center grants. Center grants, which are successfully sought and won frequently at many other peer universities, generally provide funding for multiple related grants, strong central support for infrastructure, seed grants, and administrative support. The lack of this support means missed opportunities to foster collaborative research to jumpstart ideas and plans for project grants would allow genome biology to catch-up in this area.
- 6. Unrealized links to Weill Medical School.** We are missing opportunities for well-funded and impactful science that lies at the interface of our medical school and the Ithaca campus. The two inter-campus NIH P50 grants (those of Paula Cohen and Claudia Fischbach-Teschl) are two of what could be many more such successful efforts. Biomedical Engineering is starting to realize cross-campus opportunities with Cornell's new Tech campus, but Genome Biology could easily become another major avenue of inter-campus collaboration. Universities that have a tighter connection between medical and basic sciences often reap the benefits with generous NIH funding support. There is

need for creative thinking about mechanisms to foster these collaborative efforts, starting with identification of suitable/enthusiastic collaborators on both campuses.

- 7. Uncompetitive support for outstanding graduate students and postdocs.** Other universities have elite postdoctoral programs for recruiting stellar postdoctoral fellows (e.g. The Broad Fellows program at the Broad Institute, Berkeley's Miller Fellowships or the Whitehead Fellows program). These individuals in such interactive programs can often stimulate science across labs and campuses. Cornell's lack of such positions make us less competitive than peer institutions in recruiting outstanding postdoctoral fellows. The new Presidential Postdoctoral Fellows program will certainly benefit Genome Biology at Cornell. . In addition, support is needed for graduate programs at both the Weill and Ithaca campus. These programs are able to attract applications from the very best undergraduates in the country, but too often we lose the top talent to schools that have more generous support for graduate students. Bringing in just 10 or 20 additional outstanding graduate students each year would be well worth the cost, and these students often raise the standards of students around them.

Recommended Areas of Future Growth in Genome Biology at Cornell

The Genome Biology Task Force will facilitate the recruitment of new genomics faculty to departments and colleges across campus. The individuals hired, and the focus of their research, will satisfy these criteria:

- 1) they must have a broad impact, address fundamentally important challenges for the twenty-first century. We are looking for game-changing hires.
- 2) they must build upon, and significantly enrich existing areas of strength at Cornell in genome biology, and synergize with interdisciplinary development of cutting-edge technologies.
- 3) there should exist opportunities for robust external funding.
- 4) they should be linked to the identification of outstanding faculty candidates for recruitment targeting. Such individuals would complement and be synergistic with one another as well as with the current existing strengths on the Ithaca and NYC campuses.
- 5) they should interface with other Cornell initiatives being undertaken in parallel, including Data Science, Sustainability, NEXT, and Infection Biology.

With these criteria in mind, here are our proposed areas of emphasis:

1. Genome architecture and regulation. Genomes are not inert instruction books, but instead are dynamically responsive to the needs of the cell and the organism. Those needs include changes necessitated by development to generate new cell types and functions, as well as responses to the environment, including stresses. Solving the mysteries of how the fantastically complex genome can be appropriately regulated is the driving mission of a large number of research scientists at Cornell.

We have strengths on campus in the area of genome structure and architecture that includes a funded 4-D Nucleome technology development component to an interdisciplinary and cross-college team (Co-PIs: Lin, Lis, & Zipfel, and Co-Is: Danko and Ozer). Another component of strength in this area is the study of nuclear lamins and mechanical strength and signaling by Lammerding. In the area of enhancer function, we have support by NIH RO1 grants, and also by the ENCODE (Co-PIs: Lis and Yu). Furthermore, the addition of Feschotte, who is also the recipient of a new collaborative grant from ENCODE, the NIH program for identifying and understanding the function all genomic regulatory elements (<https://www.genome.gov/10005107/encode-project/>). These collaborative efforts bring an innovative view of the evolutionary dynamics of cis-regulatory elements.

A major emphasis of NIH is to define the genome, its structural and functional units, and the regulatory interactions that govern its activity. For example, NIH is providing support the architecture of genome in the 4D Nucleome consortium project (<http://dcic.4dnucleome.org/>), which uses chromatin capture methods (like Hi-C) to define structural domains and long range regulatory interactions of the genomes. This 4D Nucleome effort is further complemented by the ENCODE project, mentioned above.

Another key aspect of genome regulation is the modulation of chromatin structure through epigenetics. Much of the difference in developmental fates of cells during development hinges on epigenetic changes, such that errors in the reading and writing of these signals is of critical importance. This is an exciting and fast-moving area, and integration of multiple “omics” techniques has allowed us to assign chromatin states to every gene with ever increasing accuracy. This in turn will present better prediction of genome responses to multiple stimuli. Many human disorders have failures in epigenetic signals, and through this understanding comes ideas for therapeutic intervention. Strengths at Cornell include Soloway, Richards, Lin, Coonrod, Rubin, Lipkin, and Melnick, who study DNA methylation and epigenetics. We have some emphasis on chromatin structure from Lee and Lis; however, a key hire here would be critical.

Much of the control of the genome is mediated through processing of RNA transcripts, both through control of alternative RNA splicing and through the regulation of RNA stability. Thousands of human genetic defects are caused by failures of accurate splicing, and already there are applications of oligonucleotide-directed splicing to correct these somatic defects. These events are often tightly coupled at the level of the whole genome, and understanding how this works is the inspiration of several investigators at Cornell, including Pleiss, Grimson, and the recent addition of Sethupathy. Extending from this area, there is a growing interest in the role of non-coding RNAs in health and disease, involving these same investigators, with the addition of Cohen, Schimenti, Rudd, and others on the WCMC campus.

2. Genome evolution and gene-environment interactions. Evolution is a critical unifying attribute of all life, and genome biology has opened doors to unprecedented knowledge of the details of how evolution has produced the life forms that exist on earth today. The study of evolution through genome sequences engages many disciplines, from population genetics to

computational biology, and essentially all aspects of genome biology have been informed by evolutionary thinking and approaches. For instance, in human medical genetics, the most widely used approach for identifying genes associated with complex diseases (GWAS) makes full use of population genetics methods. Cornell has a large and vibrant group of faculty in this area (see www.3cpg.cornell.edu) including Aquadro, Barbash, Bogdanove, Clark, Feschotte, Keinnan, Lazzaro, McCouch, Messer, Reed, Williams, Wolfner, and Yu, among many others. On-campus expertise on environmental effects on genome expression include Bogdanove, Buckler, Clark, Kwak, Lazzaro, Lis, and Sevier. In addition to its use in medical genetics, the mouse has been a valuable model for understanding evolutionary genomics in vertebrates, and Cornell boasts a sterling faculty in this area as well (Schimenti, Sethupathy, Soloway, Stover, Coonrod and Libert, among others). In the Vet school, projects such as the dog genome project (Boyko) and the companion animal biobank also make extensive use of evolutionary genomics methods.

3. Computational genomics. Already the rate limiting step for a substantial segment of life science research is computation. While some of this entails routine technical tasks like mapping millions of sequence reads to a reference genome, much of the opportunity for novel research hinges on creative applications of computer analysis. As the technologies for generating data sets that query attributes of every nucleotide in the genome expand in scope and complexity, we are facing a deluge of data that can only be solved through carefully considered application of computation. Rapid advances in machine learning, with approaches like Deep Learning, which implement unsupervised hierarchical feature extraction and multiple layers of abstraction, mean that it will soon be possible for the computer to make predictions about experimental results that far outstrip our ability to understand how the inferences were made. There is enormous scope for critically important breakthroughs at this interface of Big Data and Machine Learning, especially when it comes to making sure that the results carry meaning to us humans. On-campus faculty working in this area include Brito, Clark, Keinan, Koren, Kuceyeski, Messer, Myers, Varner, Wells, Williams, and Yu. Cornell has a stellar Computer Science department, but inadequate connections between those faculty and the exciting research projects in genomics. We have a good foundational group in computational genomics, but relative to competitor institutions, we are far behind in faculty numbers in this critical area. This is also an area ripe for expanded links with Weill Medical College. We already have a viable, shared graduate program in Computational Biology and Medicine, and there are abundant opportunities to scale up inter-campus collaborative research in computational genomics.

4. Technology as a driver of Genome Biology. New technologies have emerged from mapping nascent transcription and defining regulatory elements. These are continuing to be developed and are providing highly sensitive base-pair resolution views of gene expression at transcription. The technologies include GRO-seq, PRO-seq, GRO- and PRO-cap, and a new variant of PRO-seq called ChRO-seq that can be applied to normal and diseased tissues directly. Existing campus expertise include Kwak, Danko, Clark, and Lis. The corresponding computational development is being driven by Danko and Yu.

Rapid advances in sequencing technologies have enabled the identification of tens of millions of variants in the human population, most of which lie in the noncoding regions. Several initial studies have already demonstrated that noncoding variants can modulate gene expression and thus lead to cancer and other disorders. However, little is known about the functional impact of the vast majority of these variants. This marked lack of functional insight has become a major bottleneck for the large number of ongoing whole-genome and whole-exome sequencing projects and severely hinders our ability to convert this deluge of genomic information into useful biological or therapeutic applications. New technologies, such as Clone-seq, eSTARR-seq, PB-seq, have been developed (Yu and Lis labs) to experimentally examine tens of thousands of coding and noncoding mutations, enabling computational models to be built for prioritizing all variants.

Tracking the genome biology of single cells provides information and insights that are obscured in studies of populations of cells. Having all genome-wide technologies be performed at the single cell level would greatly stimulate Cornell's role in genome biology. Expertise in this area include Kwak and De Vlaminc.

Understanding of the molecular mechanisms of genome regulation requires defining the interactions between the genome's DNA and the protein and RNA regulators. Toward that end, chemical approaches are being developed such as bivalent chemical crosslinkers that join DNA to DNA, DNA (or RNA) to protein, and protein to protein that are photoactivatable and have defined linker lengths. This technology when coupled to genome-wide nucleic acid analysis and sophisticated mass spec methods promises to revolutionize our defining of interactions. Local expertise include Lin, Smolka, Zipfel, Danko, Yu, and Lis.

Cornell researchers have also become early adopters of genome editing technologies, most notably CRISPR/Cas9-mediated genome editing in whole organisms as well as in cell lines. Our Center for Vertebrate Genomics-based transgenic facility has generated hundreds of new mouse lines through this technology, both for the Ithaca and NYC campuses of Cornell, as well as for many outside institutions. John Schimenti has played a major role in these endeavors, along with several other investigators. Adam Bogdanove developed the first truly modular DNA editing reagents, TAL effector nucleases (TALENs), and for their high specificity TAL effector-based reagents remain an important part of the genome editing and systems biology toolbox, particularly for therapeutic application. The new Plant Transformation Facility in the School of Integrative Plant Science, directed by Matthew Willman, is driving discovery and innovation by providing full-service plant transformation and plant genome editing services for key crop species. A critical component of the genome biology initiative will be support for these transgenic facilities that provide and develop technologies that drive animal and plant genome biology. Notably, however, while we use CRISPR/Cas9 and TAL effector-based DNA targeting technology actively for our research, only part of a single lab (Ke) is studying mechanisms of DNA targeting and DNA repair with the potential to further advance this technology. A senior hire in this area would bolster our impact tremendously.

5. Establishing connections between basic regulatory mechanisms of gene expression and disease states. Cantley, Blenis, Lipkin, Nanus, Crystal, Melnick, Rubin, Elemento, and Ross provide expertise in this area and have in place a \$46M grant as part of the Personalized Genomics Initiative to collect patient samples for analysis. The full utilization of these samples would benefit from strong collaborations using the latest genome-wide experimental and computational approaches. Likewise, the cancer biology programs in Ithaca and Weill (e.g. the Physical Sciences Oncology Center and the MSKCC-Cornell Center for Translation of Cancer; including the laboratories of Fischbach, Wiesner, Cantley, Cerione, Richards, and Weiss) have the potential to benefit from the new perspective afforded by experimental and computational approaches devised by genome biologists. A new section within the Department of Biomedical Sciences is applying animal models to human diseases (Kotlikoff, Libert, Cummings, Sethupathy), a program with excellent chances for establishing collaborations with Weill investigators. DNA repair is central to our understanding of many diseases, including cancer, and that Cornell is traditionally very strong in this area (Alani, Weiss, Cohen, Schimenti, Pawlowski). Finally, the power of stem cell biology needs to be integrated into the genome biology mission, as it can provide critical insight to genome regulation and the specification of cell types during development, and it is critical in establishing therapies that involve reprogramming of cell fates.

Specific Recommendations

- 1. Shore-up existing areas of strength.** Strong disciplines like Gene Regulation, DNA Replication and Repair, and Evolutionary Genomics do not stay strong without continuing attention to maintenance of that excellence. This means not only providing competitive retention packages but also new hires in these areas. They will not stay strong without this vital renewal. One particular area that is a strength, but is seriously undersized for a university of this caliber, is Computational Genomics. We have produced a series of stars in this area, only to have them hired away. Formation of a new department of Computational Biology (CB) is a step in the right direction, and it is important that this unit grow quickly and become stabilized with a strong chair hire. We again emphasize that the success of Computational Biology at Cornell is critical to all our efforts in the broader area of Genome Biology.
- 2. Judicious hires in key emerging areas of genome biology.** Genome Biology is a fast-moving area, and vital tools and technologies need to be rapidly brought on board. Often this requires hiring new faculty with the requisite expertise. Areas such as genome editing and cryo-EM are two examples. We recommend university-wide recruitment strategies that build an integrated approach that recognizes and implements hiring in fast moving areas. A pool of start-up funds needs to be generated to facilitate rapid assembly of hiring packages. The size of the core GB Initiative remains at the level of 10 senior hires. Faculty excitement over hires in new emerging areas should be tapped by the Development Office to help secure funding for these efforts. We have already made an outstanding senior hire (Cedric Feschotte) that was influenced by the GB initiative, and this demonstrates that Cornell can be effective in recruiting in this area

at the senior level. However, we need to identify departments that will be willing to provide lines. The target departments include MB&G, Molecular Medicine, BSCB, Biomedical Sciences, Ecology and Evolutionary Biology, Chemistry, Physics, BioMedical Engineering, Integrative Plant Sciences, Nutritional Sciences, and Neurobiology & Behavior (perhaps as a joint appointment in Neurobiology and MB&G as a strategy for enhancing the genetic and molecular component of neurobiology on campus). We should consider one senior hire as someone using Cryo-EM to study a molecular machine that acts on the biochemistry of the Genome (DNA or RNA polymerase, DNA repair, etc). This would complement structural biology efforts to build a critical mass of research for developing Cryo-EM on campus. Another high priority senior hire is in the area of genome editing/CRISPR technology and its application. Finally, this effort will be facilitated by a Distinguished Genome Biology Seminar Series. We will recommend candidates for various seminar series on campus, as named lecturers (Genome Biology), with an appropriate honorarium, and that we would have them stay into a second day to have lunch with our committee. We would bring them in with the idea being that they will represent exciting/emerging areas of genome biology and that we would like to get their thoughts about our initiative.

- 3. Create infrastructure for training in genomics and in the dispersal of genomic technologies across campuses.** Incorporation of genomics technologies into the broad array of biological studies at Cornell could provide them with a major re-invigoration and a greater potential for further external funding. To get there, we need first to ensure that centralized genomic services such as DNA sequencing remain state of the art and have an extremely rapid turn-around time. Our recruiting will be very dependent on the capacity and quality of our DNA sequencing facility. Additionally, we need a training infrastructure to rapidly and effectively spread the knowledge and hands-on skills of genome biology to a broader base. We recommend a centralized structure that would be home to the latest bioinformatics and genomic molecular biology, and which would be flexible and constantly evolving. The idea is that such an infrastructure would have two-four research associates with combined experience in bioinformatics and genomic molecular biology, to actively disseminate cutting edge developments in bioinformatics and molecular genomic technologies to the broad research community at Cornell. This facility would be led by a Senior Research Associate who would provide intellectual input for grants and manuscripts, core development for center proposals, and who would spearhead/guide the Exploratory Grant mechanism (see next point). This infrastructure could be integrated seamlessly with the BRC core facilities, using their sequencing capabilities, and in turn providing new revenue-generating technologies to the BRC cores (following appropriate technology development). A structure such as this already exists under the auspices of Jen Grenier's Gene Regulation and Epigenetics Core (formerly the RNA sequencing core), which currently supports two of the three existing P50 grants on campus.
- 4. Create a genomics technology grant system to fund development and application of cutting edge genomic technologies.** We propose a "Genomics Exploratory Grant"

funding mechanism directed specifically to genome technology development that will stimulate local efforts to develop new tools in genomics. We also propose tightly integrating the technology breakthroughs with proposed mechanisms of technology dispersal (point 3). We envision technology development grants of \$50-100k where up to \$50k could be used for supplies and personnel and up to \$50k is in the form of sequencing or mass spec or other services by our core facilities. We also envision seed grants to use these new data generation technologies to provide preliminary data for external grant applications. The successful seed grant program of the Center for Vertebrate Genomics is an excellent model. We also support the small seed grants mechanism provided by the BRC, and we envision these will enable investigators to obtain the sequencing and other data needed to support extramural grant applications in genome biology.

5. Administrative support for center grants to develop genome technologies.

Investigators in the life sciences at Cornell pride themselves in their individual resourcefulness, but often fail to see the advantages of larger center grants. Universities where center grants are successfully sought and won generally provide strong central support for the administrative overhead. Further encouragement and support for seed grants to foster collaborative research to jumpstart ideas and plans for project grants would allow genome biology to catch-up in this area. (Note that the OVPR has this support at a local scale in the form of Kim Holloway, whose job is to bring faculty groups together and facilitate center grants. It may be possible to take advantage of currently existing infrastructures for administrative support of center grants, e.g., Nanobiotechnology Center. Additional infrastructure for large grant application mechanisms beyond this existing infrastructure may also be required and we recommend additional funds be committed.

6. Establish realistic collaborative links between Ithaca and the Weill Medical College.

We are missing opportunities for well-funded and impactful science that lies at the interface of our medical school and the Ithaca campus. Joint pilot grants funded from both campuses are needed for inter-campus projects that have the ability to recruit patients with diseases of interest to Cornell faculty. Biomed Engineering is starting to realize the opportunity of cross-campus coordination with the new Tech campus, but Genome Biology could easily become another major avenue of inter-campus collaboration. The new department that will focus on aspects of human genetics at Weill presents one opportunity to nucleate these collaborative efforts. Medical school investigators would often benefit by working with investigators with expertise in model organism research at the Ithaca campus. Ready access to patient samples would be a big boost for many investigators on the main campus. The Precision Medicine Initiative at Weill could also be the focus of many inter-campus collaborations, especially in the area of computational biology and patient recruitment core services for Cornell faculty. An ongoing example of a joint training program is the Tri-Institutional Program in Medical Computational Biology, and a previous example is a Keck Foundation grant shared between Cornell's Ithaca and medical college campuses. To promote new pilot projects

that have the potential for a similarly large impact, we propose seed grants that would fund genome biology projects specifically between collaborators in Ithaca and NYC. We look forward to working together with Dean Gary Koretzky's and his program to fund collaborative projects between the campuses.

- 7. Better support for graduate and post-graduate training.** The ability to attract the best graduate students and postdocs into our research programs is hands-down the most critical rate-limiting step to achieve excellence in genome biology. Cornell is able to attract strong students and postdocs, but this is an extremely competitive area, and it takes constant effort (and money) to maintain this recruitment. We support the University's existing efforts to strengthen university-wide scholarships at both graduate and postdoc levels. These would be well advertised, highly prestigious and competitive, similar to the Miller postdoctoral fellowships at Berkeley, the Broad Fellows program at the Broad Institute, the Whitehead Fellows and Harvard Fellows programs. In order to emphasize the theme of Radical Collaboration, we propose that the fellows must identify a pair of advisors to work with, and define a collaborative project that spans the work of the respective labs. The fellows would also have a sense of community, with meetings and shared mentoring. Thus, we recommend a commitment to a Fellows Program in Genome Biology that would be in addition (or be an augmentation) to the broader Cornell Fellows Program.

Timeline for Implementation

1. Implement a seminar program to bring in outstanding speakers in Genome Biology. The speakers will be potential candidates and serve as external consultants. The goal is to bring in 12 speakers per year. This has been initiated and will be in full implementation in Fall of 2018.
2. With the goal of enhancing genome biology on campus and recruiting outstanding individuals, expand support to improve sequencing facility immediately (before end of year 2017) and recruiting two research associates to facilitate spreading of technology across campus (by summer 2018). Utilization of these infrastructures and dissemination of these methodologies will be augmented by the availability of a seed grant program (by summer of 2018).
3. Hire two outstanding senior genome biologists a year for the next 4 years with time line beginning in Spring 2019.

Appendix A

Genome Biology Task Force Charge:

The Genome Biology Task Force will develop a strategic plan that includes recommendations to shape future recruitment, graduate training, and infrastructure investments for the coming years. While the plan will shape investments on the Ithaca campus, the task force includes key participants from Weill Cornell Medicine in order to further unify our campuses academically and take advantage of aggregate strengths in genomics and genomic medicine in developing multi-investigator proposals. The task force will shape and strengthen programs in gene regulation, genomics, gene repair, and genetic medicine, as well as other areas of basic and applied biology. The planning exercise should be completed within the 2016-17 academic year, and it is anticipated that the task force will play a critical, ongoing role in coordinating and facilitating recruitments, fostering collaborations, and implementing other aspects of the recommendations.

Membership:

Name	Affiliation
Apostolou, Effie	Weill Cornell Medicine, Molecular Biology in Medicine
Baird, Barbara	Chemistry and Chemical Biology
Bogdanove, Adam	Plant Pathology and Plant-Microbe Biology Section
Cantley, Lew	Weill Cornell, Department of Medicine
Cerione, Rick	Chemistry and Chemical Biology
Clark, Andrew	Molecular Biology and Genetics
Cohen, Paula	Biomedical Sciences
Feschotte, Cedric	Molecular Biology and Genetics
Fischbach, Claudia	Biomedical Engineering
Koretzky, Gary	Vice Provost for Academic Integration
Lipkin, Steve	Weill Cornell, Department of Medicine
Lis, John	Molecular Biology and Genetics
Sethupathy, Praveen	Biomedical Sciences
Soloway, Paul	Biomedical Sciences
Yu, Haiyuan	Biological Statistics and Computational Biology