FINAL

National Human Genome Research Institute
Planning Meeting Summary

Genomics of Gene Regulation, October 27-28, 2009

Purpose and Introduction

One of the great challenges in the interpretation of the "book of life" is the comprehensive identification of the regulatory sequences determining where and when every protein-coding gene, and noncoding gene, is switched on and off throughout the human life cycle. Stated as a strong hypothesis, it should be possible to infer gene activity solely from primary sequence—the "language of gene regulation." With adequate understanding of the rules under which regulation operates, we should be able to look at DNA sequence and predict where, when, and how much any gene will be expressed.

This has particular relevance to understanding human disease: Over half of all genetic variants conferring risk to human disease reside in noncoding, probably regulatory, sequences. A full understanding of the language of gene regulation will empower us to predict abnormal function in cases of inherited disorders. It will also enable, and go along with, much deeper understanding of basic biology such as animal development.

On October 27-28, 2009 the National Human Genome Research Institute (NHGRI) convened a workshop to discuss this broad area of research. How close can we get as a scientific community to the goal of complete understanding of the rules of gene regulation? What more will we need to learn in order to get there? What genomic approaches to this challenge are both valuable and feasible over the next 5 to 10 years? What will it take to make them feasible?

Participants were asked to discuss these issues from three viewpoints: the function of primary regulatory DNA sequences; natural and experimental variation of gene regulation; and gene regulatory networks.

This workshop was deliberately broader in scope than one NHGRI would convene to plan for a specific initiative, seeking ideas that were significant and timely, but not (for the present) constricted by timeframes or budgets.

The outcome of this workshop will be used to aid long-range scientific planning activities at NHGRI. This workshop was also held in the context of a larger NHGRI scientific planning effort, and its outcome will be evaluated in that larger context.

Summary of Final Recommendations

Due to the breadth of discussions, the final recommendations covered multiple areas of research and development. The recommendations below are listed in no particular order of priority.

Participants called for the following:

Basic knowledge:

- **Perform comprehensive mechanistic analysis of a representative set of cis-regulatory elements.** We have only a handful of examples of genes for which cis-regulation is understood to a very high level of detail, including understanding cis-regulatory sequences (eg. by systematic mutation), and the biochemical details of factor binding to those sequences. It is also critical to understand how factors, which are bound to cis elements, act through the help of coregulatory proteins to influence transcription and transcription-coupled RNA processing. While it is not feasible to obtain this level of detail for all genes, it may be possible, for a relatively small amount of funding, to find a small number (20? 50?) of examples that could provide a much better basis for generalization to all cases. This data set would, among other things, inform high throughput functional tests for DNA and RNA elements, as well as chromatin structure/dynamics.

- **Obtain more detailed information about the structure and, in the longer term, dynamics of chromatin and how it relates to expression (transcription and accessory factor interaction, chromatin looping, etc.).** This is an area of very active ongoing study. Much more detailed knowledge will be required, and will be done. New methods and reagents could accelerate this area of research (see below).

- **Obtain additional validated example gene regulatory networks.** We do not yet have enough high-quality examples from either development or physiology.

- **Extend gene network analysis to phenotype via downstream network components**. Current gene network analysis is most detailed in describing parts of networks that are remote from their actual expression as phenotypes. Portions of regulatory networks that are more proximate to anatomical or physiological phenotypes need to be elucidated.

- **Undertake network level phylogenetic comparisons** and synthetic network design.

Technologies/Methods/Reagents:

- **High quality, readily available reagents will accelerate the field.** Monoclonal antibodies and other affinity reagents for all regulatory factors should be developed.

- **Methods to obtain information about DNA binding need ongoing development.** For example, improvements in ChIP or, in the longer term, a better alternative.

- **Develop improved methods for genomic manipulation** both as a means to understand DNA elements and also to understand gene networks (synthetic network design).

- **Develop methods for single cell (or smaller amount of material) analysis (including imaging).** Single cell-analysis is important, for example, to overcome problems associated with measuring the relationship between factor binding and expression as combined information from large numbers of cells.

- **Develop experimental models** for understanding natural or disease-associated variation in a regulatory circuit.

Resources/Projects:

- **For human studies, undertake a massive cohort study** to provide the basis for studies elucidating the contribution of variants in regulatory DNA (including UTRs) to phenotype and to identify variation in regulatory factors. This would be integrated with variation and functional data. This would clearly be a very large, long-term project.

- **Establish and carry out high-throughput functional tests for DNA and RNA elements** that control gene expression (in many cell types, developmental, and physiological states) prioritized intelligently, including perturbation methods, and in the longer term, genomic manipulations.

- **Provide access to the data** from these (& other) projects for the average researcher via unified databases (a long-term goal) and robust software tools, and to support computational modeling. Should be linked to other large-scale projects.

National Human Genome Research Institute
Planning Meeting Summary

Genomics of Gene Regulation, October 27-28, 2009

Full Report

How close can we get as a scientific community to the goal of complete understanding of the rules of gene regulation? What more will we need to learn in order to get there? What genomic approaches to this challenge are both valuable and feasible over the next 5 to 10 years? What will it take to make them feasible?

This workshop report summarizes a discussion of these major questions convened by the National Human Genome Research Institute (NHGRI). A full description of the aims of the workshop, and a workshop agenda, are in Appendix 1.

This report will not describe all presentations in detail, but will rather focus on the major conclusions in an attempt to extract the most important points related to the objectives of the workshop.

I. Transcriptional Regulation: Some General Lessons

Three plenary speakers (John Lis, Kevin Struhl, Gary Felsenfeld) reviewed the high-level conclusions about our current knowledge of transcriptional regulation that are relevant to this workshop.

1. *Characterization of transcriptional regulation must account at least for cis-regulatory sequences, the regulatory factors that bind them, co-regulatory mediators, polymerase (PolII), and the interactions between all of these elements. We also need to understand the rate-limiting steps in transcriptional activation, any of which could be possible targets for regulation by the cell.*

   Investigation of transcriptional control at heat shock loci in *Drosophila melanogaster* provides an illustration at a high level of detail. Studies locating factors that bind to cis-regulatory sequences (and extending to understanding the dynamics of their distribution), using techniques such as immunoflourescence and ChIP, have built up the view that PolII is primed and actively (i.e., requiring stabilization by other factors) paused at the core promoter. Activation is accomplished by successive binding and recruitment of a variety of cis-regulatory factors.

   In order to understand transcriptional activation better, for example, to be able to distinguish which rate limiting steps are actually most relevant for regulation, we will need methods that allow the probing of dynamic aspects of factor binding. Ideally, we also need methods that can be applied across the genome, for example, to examine the dynamics of factor binding and PolII activation in

response to a stimulus. Some methods already exist. For example, recent genome-wide data on engaged PolII points to the conclusion that escape from pause may be more significant than the initial engagement of PolII.

In addition to methods that dynamically image the major components involved in transcriptional activation, we need improved methods to perturb---and so test--their function.

2. *To understand transcriptional regulation, we need to relate the details of how, where, and when proteins bind to specific sites on DNA in vivo, to the dynamic role of the components of chromatin. One lesson of current knowledge is that we must be wary about making predictions based on DNA sequence motifs alone. Neither can we completely rely on invoking chromatin structure to understand how transcriptional states are determined.*

   The importance of understanding the relationships between DNA binding and chromatin are obvious in considering that genetically identical cells can inherit different on/off states of gene transcription (epigenetics), a ubiquitous feature of metazoan development.

   How do short DNA sequences actually specify binding? The answer into this question is complicated by the observation that genomes appear to be very permissive to the general binding of factors, and that there is a continuum of binding *in vivo.* On top of that, there is not necessarily a straight correlation between strength of binding and activity. Moreover, simply because a good motif exists, does not mean that it is bound—for example, most of the p63 binding motifs in human DNA are not occupied by p63. Moreover, many "poor" p63 binding sites are occupied. DNA sequence alone contributes to, but is a poor predictor of, DNA binding *in vivo.*

   In addition to understanding binding specificity, there are other circumstances that determine whether factors actually bind, including interactions with the many other factors that constitute chromatin, accessibility of the DNA, local structure of DNA and chromatin, and others. For example, nucleosome positioning can result in a state accessible to binding. So, chromatin structure can facilitate binding, but does not determine it. There are many examples of the inter-relationship between the two that support the idea that neither is sufficient to understand transcription.

   So, in addition to understanding binding specificity based on sequence, we will also need to account for how chromatin mediates accessibility and cooperatively between transcription factors, and how the interactions between binding and chromatin dynamics give rise to transcription. All of these complications must be considered in defining a functional binding site.

3. *In order to understand transcriptional regulation, we need to take into account the long-range interactions that occur between regulatory sequences that may be*

distant from each other, at several scales. These physical networks of interacting sites within the nucleus are maintained by a wide variety of specific interactions. Some of these structures may be stabilized by site-specific DNA binding proteins dedicated to that task. Many of these interactions may be important for regulation of gene expression.

At a small scale, there is considerable evidence that adjacent gene systems interact, and we know some of the sequence elements that mediate this. For example, enhancers can act at a distance on multiple nearby genes, and will do so unless blocked by an insulator. Insulators themselves can be regulated (eg, by CTCF binding). The mechanism by which this occurs may entail CTCF binding with itself or with a surface (e.g. nucleolar surface) to result in loops or other features that contribute to (and correspond with) both long-range interactions and the higher-order conformation of the chromosome within the nucleus. Multiple molecular mechanisms are likely to underlie looping.

There is growing evidence for the existence, and importance of, specific longer-range interactions. Microscopy provides evidence for chromosome domains, and for the idea that loci that are regulated by the same transcription factors co-localize in the nucleus. Detailed cross-linking/sequencing studies (4C and higher order) capture regions that are associated within the nucleus, across the entire genome at increasingly high resolution. There is a need for additional genome-wide methods that will both increase resolution and allow better correlation with expression.

II. Breakout Session Summaries

The major points from the first day of talks were discussed in three breakout sessions, each of which was intended to address a different aspect of the genomics of gene regulation: Primary Regulatory DNA Sequences: Function, Analysis, and Prediction; Natural and Experimental Variation of Gene Expression; and Gene Regulatory networks.

a. Primary Regulatory DNA Sequences: Function, Analysis, and Prediction

This topic is intended to investigate genomic approaches to characterizing and understanding the range of functional sequence elements that control gene expression (e.g., enhancers, silencers, insulators, core promoters, tethering elements, elements mediating long-range interactions, control via epigenetic modifications, etc.).

The breakout group identified three inter-related questions that encompass this topic:

- What are all the elements that control gene expression? e.g., functional assays for enhancers, silencers, etc. in all cell types
- What is the complete structure of chromatin? e.g., transcription factors, accessory factors, looping in all cell types

- What are the key regulators (and sequences) of cell state? e.g. forward genetics, reprogramming, etc. in key cells

The group proposed multiple research and resource development efforts and general guidance that could address these questions, divided into the following categories: Organisms; Systems; Cell Types; Data Types; Technologies and Reagents; "Blue Sky" Ideas. These suggestions were made keeping in mind the charge.

- Organisms used for systematic study of this problem should include:
  - Human: most challenging, but has the most clinical relevance
  - Well-characterized vertebrate systems (mouse, zebrafish)
  - Select model systems to provide a more rapid path to understanding basic principles

- Systems to be studied should focus on:
  - Early development, where there is considerable existing knowledge; early developmental pathways are often also associated with disease
  - Systems that are also amenable to predictive, testable modeling

- Cell types. In order to make systematic progress, it will be necessary to focus on a finite number. The following were mentioned due to multiple characteristics, including previously accumulated knowledge, ease of manipulation, accessibility, disease relevance, etc.:
  - Embryonic
  - ES and iPS cells
  - Hematopoietic stem cells
  - Somatic cells that can easily be obtained in large quantities (T and B cells)
  - Newly isolated tumor and tumor stem cells

- Data types that could be productively pursued in the next five years (suggested for three model organisms and three vertebrate cell types) include:
  - ChIP-Seq of ~2000 TFs, accessory factors, nucleosomes
  - Protein-DNA binding motifs (in vitro)
  - Protein-protein interactions (for proteins related directly to transcription)
  - RNA expression (RNA-Seq, GRO-Seq)
  - Expression following shRNA knockdown of regulators
  - Enhancer/silencer element reporters
  - Hi-C looping assays
  - Forward genetics
  - Cellular reprogramming

- Data types that may be possible to obtain in the next ten years include:
  - High throughput element knockout
  - Protein-DNA structure (atomic resolution)
  - Protein modifications
  - Advanced looping/compaction assays

    - RNA elements
    - Imaging
    - Small molecules

- Technologies and reagents that should be sought include:
    - Computational methods and models
    - Methods to connect elements to target genes
    - Advanced chromosome structure analysis
    - Single cell and minimal cell analysis
    - ChIP-grade antibodies
    - Tagged factors
    - All regulators (inducible) in vectors for reprogramming
    - ChIP replacement (a better technology for assaying protein binding to nucleic acids)

Finally, the group outlined three "Blue Sky" ideas. These were not meant to be feasible in any limited timeframe, but rather were intended to stimulate ambitious long-term thinking:

- For one diploid human: characterize all cis-regulatory elements in all cells
- Identify all variant elements in iPS cells from patients encompassing all diseases
- Develop small molecule agonists and antagonists for regulatory elements involved in all diseases

b. Natural and Experimental Variation of Gene Expression

This breakout session was intended to bring out what has been, and could be, learned about the genomics of gene regulation from studies of variation. What do laboratory experiments, natural variation, and evolutionary changes tell us about the language of gene expression? What are the current lessons from comparative genomics; studies of natural variation; recurring variation underlying recurrent phenotypic changes; studies of regulatory variation (eQTLs linked to anatomical/physiological phenotypes); and human medical sequencing studies that uncover regulatory variation? What is our current capability to manipulate regulatory elements, and what general lessons have been learned?

The breakout group began by noting that the critical emphasis of its discussion was on understanding the relationship between inter-individual variation and disease. In the context of this workshop, how much of the segregating variation in disease risk and other medically important phenotypes is caused by polymorphisms whose primary proximal effect is on the regulation of gene expression? This gives rise to the following questions:

- How do polymorphisms that directly affect protein levels break down between transcription rate, transcript stability, transcript localization, translation rate, etc...?
- What is the spectrum of variants that affect each of these processes?

- What portion of variants (of different classes - SNPs, indels, CNVs, etc...) affect gene expression - how many have no effect on transcript/protein level?
- What is the relationship between variation in gene expression and variation in fitness (i.e. how much, and what types, of expression change is neutral)?

The group also noted a few points that were considered in their discussion:

- Rare variation can have significant phenotypic effect, and be important for functional insight, even if not for understanding common disease risk.
- We can understand a complete regulatory circuit, yet know little about the consequences of variation in genes in that circuit. Both are necessary to understand disease. We probably cannot understand the later by analyzing one variant at a time.
- A goal of variation studies is to understand the spectrum of effects of gene sequence variation within the human population that result in differences in gene regulation. We also probably cannot understand this by analyzing one variant at a time.

The working group discussion encompassed two diverging viewpoints. The first asserted that it was critical to characterize variants, probably in an unbiased way, in human populations in order to understand their contribution to gene expression, phenotype, allelic spectrum, and common disease risk. The second was very concerned about the idea that the vast majority of SNPs would have little or no effect, making any unbiased approach to studying them inefficient. As a concrete example, there was a concern that an effort to comprehensively analyze SNPs (for example through reporter assays) would be too inefficient, compared to a mutational analysis, e.g. in cell lines followed by selection. On the other hand, this approach would miss some critical aspects of understanding variation and how it relates to disease in human populations. The final recommendations emphasized the need to study variation in the context of human populations.

In order to provide the scientific basis for understanding the relationship between regulatory variation and phenotype in a way that is clinically relevant to the human population, the working group proposed the development of a foundational integrated resource. This resource would include:

- A massive cohort study to maximally leverage individual researcher-initiated studies.
- DNA sequence of many individuals within the study (10,000+), along with anonymized medical records.
- Generation of functional data from a significant subset of these individuals: RNA-seq from primary tissues, RNase I hypersensitivity (allele-specific), chIP-seq for many TFs, miRNA quantification.
- Bioinformatic infrastructure, including the ability to query a variant database, linking all these attributes to functional measures, genome-wide association, and other results.

- Biological materials (primary and iPS cell lines) to be readily available.

Such a large cohort study would have wide clinical relevance. For example, a publicly available cohort database would allow investigators to narrow a list of candidate SNPs and genes associated with dysregulation and/or disease states for follow-up functional work. Any cross-sectional sample of the human population, if large enough, would permit phenotypic screening that could identify alleles of large effect. Finally, it would permit investigators to predict functional effects of variants, so that when the next case comes into the clinic, his/her rare variant can be recognized.

It is easy to imagine how such a cohort study on its own and integrated with other studies would lead to an understanding of regulatory variation. It could result in, for example, an annotated database of attributes of human variants, linking genotypes, allele-specific expression levels, DNase I hypersensitive sites, chIP-seq and medical records/disease phenotypes, and integration with network models. Ideally, it will lead to the ability to predict the functional consequences of specific variants.

Along with such a resource, model organism studies will continue to be valuable to assess functional variation through studies of natural variation of model organisms, allow transgenic introduction of human variant proteins into model organisms, provide basic insight into the cis-regulatory variation most likely to have functional effect, and model and understand evolution of regulatory divergence among species.

The working group identified several technical obstacles to such an integrated resource:

- Highly specific monoclonal antibodies for each transcription factor.
- Faster, cheaper, better RNA-seq (and GRO-seq, translation rate assays, protein stability assays, etc.)
- Faster, cheaper, better methods for testing functional effects of variation in regulatory sequences.
- Faster, cheaper, better ways to measure environmental exposures.
- Ability to store and retrieve dozens of tissues of thousands of samples.
- Bioinformatic integration from genome-wide association to functional studies.

The working group noted several conceptual issues that would need to be resolved in order to make best use of such a resource:

- Does it makes sense to think of the regulatory effects of one SNP at a time – or is context always critical?
- We need to better understand how variation at the sequence level mediates variation at the gene regulatory level; we need quantitative, predictive models for this.
- We need statistical tools to deal with the high dimensionality of the primary data, and a better understanding of the error properties of the measurements (*e.g*. RNA-seq).

c. Gene Regulatory Networks

This session was intended to explore how transcriptional regulation is coordinated among genes that act together to produce a phenotype. Can we go beyond transcriptional "wiring diagrams" to make predictive network models? Can we go beyond correlating expression with individual cell states to dynamics (e.g., perturbations, developmental biology)? What is the most important/meaningful/useful type of information to obtain in characterizing a gene regulatory network? What are the most promising experimental and analytical approaches, and how would they inform an understanding of the language of gene expression? For this topic in particular, what are the technical (experimental, computational, analytical) challenges to progress?

The breakout group discussion identified several fundamental objectives related to this topic:

- What are the (inter-related) structural and functional principles of cis-regulatory modules of all kinds?
- What are the system-level architectures of genomically encoded regulatory interactions that control diverse phenotypes (e.g., development of multiple taxa; different physiological responses)?
- What are the design principles of developmental and physiological regulatory networks, and how can we conceptually abstract them?
- How are the functional activities of cells (e.g., physiology, "housekeeping") causally linked to the upstream genomic control networks (e.g., those underlying differentiation)?

Much of the discussion reflected the situation that we have relatively few examples of well-characterized regulatory networks, especially in metazoans. To the extent that we do, these are usually limited to early development, and we still lack a sufficient variety of examples. Genomic approaches have begun to allow investigation of regulatory networks (e.g., through large-scale expression experiments and perturbation, along with analysis capabilities), but these are not sufficiently developed. Most of the group's specific recommendations reflected the idea that we have just begun to investigate this topic, that the problem is large in scope (multiple genes and cis-regulatory relationships, multiple cells/cell types/tissues/developmental time points, multiple organisms, etc). That is, the recommendations focused the need for developing better tools to obtain the relevant data and resources to support further advances, initially by providing more examples of validated, biologically relevant genome regulatory networks.

It was also pointed out that some thought should be given to how best to organize any concerted effort to elucidate gene networks. The problem is still formidable. For a larger project, it may be useful to work collaboratively with other institutes to address problems that are biologically relevant, for example, on an organ system. Another approach would support consortia that will approach a network problem from different directions (e.g. Hox genes, embryonic systems, organogenesis, and physiological response). Either would provide critical, high-quality network examples, guide the project towards the

appropriate samples and model systems, and encourage biological validation of network models, while also providing a coordinated mechanism to develop relevant generalizable research tools.

At the same time, the working group noted that a lot can be learned from mid-scale projects (e.g., dealing with 100's of genes), and that as technological capabilities are improved and community resources developed, more will become possible at smaller scale, and that in fact, much of the work would occur at this scale pursued by investigator-initiated research.

The group provided examples of high-priority, highly useful research directions as context for their specific recommendations:

- Obtain more high quality validated example networks, in a range of contexts (embryonic and postembryonic development, physiology, etc.).
- Obtain more examples of exhaustively characterized cis-regulatory modules to get mechanistic structure-function principles. Enough examples to cover the diversity of cis-regulatory design. Another approach would be an effort to exhaustively characterize a network of 20-50 genes, the factors that regulate them, chromatin modification, and their outputs, as a publicly available resource to stimulate development of analysis methods, and generally to provide basic information about how to design more comprehensive studies.
- Understand network structure and modularity by appropriate phylogenetic comparisons at network level.
- Establish the linkages from upstream to downstream network modules (physiology, differentiation, morphogenesis).

The group recommended further development of the following methods as being of high importance to overcome major bottlenecks in gene network analysis:

- Develop better perturbation methods/resources – this includes small molecules, knockouts, etc.
- Develop better capabilities for manipulating genomic regions/gene transfer/recombineering –goes beyond perturbations to testing multiple elements of cis-regulatory networks.
- Develop single-cell analysis methods for a wide range of studies (for example, including biophysical constants of factor binding, expression, advanced imaging technologies, etc.) that are accessible to smaller labs.
- Develop improved methods for cis-regulatory analysis (spatial/temporal/quantitative) in single cells, tissues, and organisms.

All of these should aim to work in a biologically relevant context.

The group also noted the need for improved analytical and informatics capabilities to deal with large, complex datasets that will be generated:

- Develop improved modeling methods to deal with, e.g., large networks to discern modularity, meta-analysis of network structure, spatial information, relating data to morphological context.
- All computational and analysis tools (e.g., inferring a network from perturbation data) should be robust enough that they can be widely used by the community.
- Develop a way to capture and curate data from individual lab efforts (including negative data) for community meta-analysis.
- "Sysbase" as a community research database/platform (across model systems, human, etc.).

III. Summary and Discussion

This workshop deliberately sought to cover a very wide range of topics relating to the genomics of gene regulation. As a result, the recommendations from individual breakout sessions were diverse, and difficult to summarize as an integrated whole. For example, there was no single project recommendation that represented the clear next step in genomic exploration of the language of gene regulation, nor was there a single consensus vision of a large, long-term genomic project that would provide a comprehensive view of the genomics of gene regulation.

Rather, there were recommendations for multiple projects, resources, and tools that would together advance our knowledge of the genomics of gene regulation substantially. Many of these recommendations would also aid other fields. In the discussions, the priority of a number of the breakout group proposals was reinforced either because they were proposed by multiple groups, or because discussion reflected overall enthusiasm for the ideas. These proposals appear below.

This multiple result appears to reflect a theme that recurred in all three breakout sessions—namely, in each major area of discussion, we lack a lot of the basic knowledge that would inform how to pursue a single systematic effort. For example, we have very few examples of well-validated, detailed cis-regulation of genes, and what we do know is not straightforward or, so far, generalizable. This means it is difficult to tell what would be most valuable to do as a large-scale effort—for example, a full characterization of regulatory binding sites may be necessary, but would not be adequate, because the relationship between the presence of a binding site and the activation of transcription is complex. Similarly, we are just beginning to explore the relationship between variation and phenotype in human populations. And we only have a few extensive, validated examples of gene regulatory networks.

Because the workshop was so broad, some thoughts were raised about how best to organize the research. For example, what is the best balance of investigator-initiated, mid-scale, and large efforts? How should specific proposals for large efforts be justified and compared? How can NHGRI work collaboratively with other institutes with a basic or applied interest in this area? How can data produced at multiple scales, in multiple venues be made available to the community? These considerations led some participants to consider that the most important directions in this area should be pursued through

investigator-initiated efforts. This tends to place the emphasis for NHGRI on development of tools, methods and resources that can aid the entire field at multiple scales, and on select, well-justified efforts that are of appropriate scope—some of them large-- to begin to examine the critical problems of the field in a systematic way.

One point raised in one of the session talks and again in the summary discussion reflected on the challenge of the sheer amount of data that could potentially be systematically collected to bear on this subject. Although it is ideal to have completely unbiased "-omic" datasets, when we consider this set of problems we are implicitly considering data on cis-regulatory regions for multiple genes, multiplied by multiple cell types, multiplied by different conditions or time points, etc. For practical reasons alone, we will need to find a way to constrain what information we seek in a way that will provide the most biologically significant information and/or seek the most relevant examples. In this regard, it is useful to consider placing the workshop discussions in the context of two high-level goals:

1. Understand the molecular function of every noncoding element that contains polymorphisms predicted to affect the function of the element.

2. Understand how the variation in each element affects the function of the genetic network in which it is embedded, and ultimately leads to phenotypic variation.

In summary, the workshop participants called for the following to be developed:

Basic knowledge:

- **Perform comprehensive mechanistic analysis of a representative set of cis-regulatory elements.** We have only a handful of examples of genes for which cis-regulation is understood to a very high level of detail, including understanding cis-regulatory sequences (e.g. by systematic mutation), and the biochemical details of factor binding to those sequences. Moreover, it is also critical to understand how factors, which are bound to cis elements, act through the help of coregulatory proteins to influence transcription and transcription-coupled RNA processing. While it is not feasible to obtain this level of detail for all genes, it may be possible, for a relatively small amount of funding, to find a small number (20? 50?) of examples that could provide a much better basis for generalization to all cases. This data set would, among other things, inform high throughput functional tests for DNA and RNA elements, as well as chromatin structure/dynamics.

- **Obtain more detailed information about the structure and, in the longer term, dynamics of chromatin and how it relates to expression (transcription and accessory factor interaction, chromatin looping, etc.).** This is an area of very active ongoing study. Much more detailed knowledge will be required, and will be done. New methods and reagents could accelerate this area of research (see below).

- **Obtain additional validated example gene regulatory networks.** We do not yet have enough high-quality examples from either development or physiology.

- **Extend gene network analysis to phenotype via downstream network components**. Current gene network analysis is most detailed in describing parts of networks that are remote from their actual expression as phenotypes. Portions of regulatory networks that are more proximate to anatomical or physiological phenotypes need to be elucidated.

- **Undertake network level phylogenetic comparisons** and synthetic network design.

Technologies/Methods/Reagents:

- **High quality, readily available reagents will accelerate the field.** Monoclonal antibodies and other affinity reagents for all regulatory factors should be developed.

- **Methods to obtain information about DNA binding need ongoing development.** For example, improvements in ChIP or, in the longer term, a better alternative.

- **Develop improved methods for genomic manipulation** both as a means to understand DNA elements and also to understand gene networks (synthetic network design).

- **Develop methods for single cell (or smaller amount of material) analysis (including imaging).** Single cell-analysis is important for example to overcome problems associated with measuring the relationship between factor binding and expression as combined information from large numbers of cells.

- **Develop experimental models** for understanding natural or disease-associated variation in a regulatory circuit.

Resources/Projects:

- **For human studies, undertake a massive cohort study** to provide the basis for studies elucidating the contribution of variants in regulatory DNA (including UTRs) to phenotype and to identify variation in regulatory factors. This would be integrated with variation and functional data. This would clearly be a very large, long-term project.

- **Establish and carry out high-throughput functional tests for DNA and RNA elements** that control gene expression (in many cell types, developmental, and physiological states) prioritized intelligently, including perturbation methods, and in the longer term, genomic manipulations.

- **Provide access to the data** from these (& other) projects for the average researcher via unified databases (a long-term goal) and robust software tools, and to support computational modeling. Should be link to other large-scale projects.

Finally, a consequence of organizing a workshop with a scope this broad is that the details of organizing a specific project, or prioritizing among alternate related projects, will be neglected. It was anticipated from the outset that, once basic directions were identified, organizing additional detailed workshops could be useful.

Genomics of Gene Regulation Planning Workshop

Gaithersburg Marriott Washingtonian Center
October 27-28, 2009

AGENDA

Preamble

One of the great challenges in the interpretation of the "book of life" is the comprehensive identification of the regulatory sequences determining where and when every protein-coding gene, and noncoding gene, is switched on and off throughout the human life cycle. Ultimately, it should be possible to infer gene activity solely from primary sequence—the "language of gene regulation."

This has particular relevance to understanding human disease: Over half of all genetic variants conferring risk to human disease reside in noncoding, probably regulatory, sequences. A full understanding of the language of gene regulation will empower us to predict abnormal function in cases of inherited disorders.

So, with adequate understanding of the rules under which regulation operates, we should be able to look at DNA sequence and predict where, when, and how much any gene will be expressed. How close to that goal can we get? What will we need to learn in order to get there? What genomic approaches to this challenge are both valuable and feasible over the next 5 to 10 years? What will it take to make them feasible? Some examples include:

- Target binding data sets for all known transcription factors across the genome, including both preferred recognition sequences and variants.
- High-throughput, comprehensive analysis of cis-regulatory DNAs. This could include systematic perturbation assays, analysis of synthetic regulatory DNAs, and analysis of sequence variants found in divergent populations.
- Currently, there is an insufficient data set of known regulatory RNAs with known in vivo activities.
- Comprehensive gene network analysis.

We hope that this workshop will aid long-range planning activities at NHGRI. Because the workshop is deliberately broader than one we would convene to plan for a specific initiative, we seek ideas that are not (for the present) excessively constricted by timeframes or budgets. We ask that ideas be developed, with arguments for their significance and timeliness.

A summary of the workshop will be pursued. That summary may take the form of one or more "white papers" outlining the best opportunities in this area for NHGRI to pursue over the next 5 or more years. Because of the open-ended nature of this workshop, significant ideas that come out of it may be the subject of more focused workshops in the future.

Day 1, October 27, 2009

8:30 a.m.

1. Welcome NHGRI Staff

8:35 a.m. -9:00 a.m.

2a. Purpose, Scope, and Boundaries: "Deliverables" Workshop Chair: Rick Myers

2b. Background: The ENCODE/modEncode Projects Rick Myers

9:00 a.m. - 10:45 a.m.

3. Plenary Talks and Discussion Moderator: Mike Levine

"Mechanisms of Transcriptional Regulation: The State of the Art and Beyond"

What we already know about the mechanisms of transcription provides the foundation for understanding the language of gene expression. Starting with examples of their own work, speakers will share their outlook on the state of the art in the field, discuss major unanswered questions, and provide some context into how the work they describe fits into understanding the larger issue of the language of gene expression.

9:00 a.m. - 9:30 a.m. John Lis

"Probing Regulatory Mechanisms In Vivo Using Technologies That Are in Hand and in Our Dreams"

9:35 a.m. - 10:05 a.m. Kevin Struhl

"How Proteins Find Their Target Sites In Vivo: Implications for Epigenetic Inheritance"

10:10 a.m. - 10:45 a.m. Gary Felsenfeld

"Effects on Gene Expression of Long-Range Interactions in the Nucleus"

11:00 a.m. - 11:15 a.m. ***Break***

4. Three Aspects of the Language of Gene Expression

For each topic below, topic chairs will discuss their work within the context of the scientific state of the art and identify major unanswered questions, challenges, etc. Each summary talk will be followed by three talks that start with the example of the investigator's work and then go on to address what we want to do/learn/discover that we can not currently and what we can do in the next 5 to 10 years to achieve that, including scientific topics to pursue, computational and technological limitations, etc. Discussion should focus on scientific goals rather than recommending specific projects. Each topic chair will have 20 minutes and each speaker 15 minutes. Each session will have time for discussion.

11:15 a.m. - 12:30 p.m.

4a. Primary Regulatory DNA Sequences: Function, Analysis, and Prediction

This topic integrates a variety of ideas and data sets needed to address this question at several functional levels, addressing the range of functional sequence elements that control gene expression (e.g., enhancers, silencers, insulators, core promoters, tethering elements, elements mediating long-range interactions, control via epigenetic modifications, etc.). There are now many detailed transcriptional regulatory "stories" featuring these elements that set the stage for thinking at the genomic scale.

"Primary Regulatory DNA Sequences: Function, Analysis, and Prediction"
Chair: Rick Young (20 min)

"Whole-Genome Analysis of Drosophila Embryogenesis"
Mike Levine (15 min)

"Lessons From ENCODE: Progress and Challenges in Determining Regulatory Element Function"
Jason Lieb (15 min)

"Regulatory Element Discovery and Characterization"
David Gifford (15 min)

12:30 p.m. - 1:30 p.m. ***Lunch***

1:30 p.m. - 2:45 p.m.

4b. Natural and Experimental Variation of Gene Regulation

What do laboratory experiments, natural variation, and evolutionary changes tell us about the language of gene expression? What are the current lessons from comparative genomics; studies of natural variation; recurring variation underlying recurrent phenotypic changes; studies of regulatory variation (eQTLs linked to anatomical/physiological phenotypes); and human medical sequencing studies that uncover regulatory variation? What is our current capability to manipulate regulatory elements, and what general lessons have been learned?

"Applications of Allele-Specific Expression From RNA-seq Data"
Chair: Andy Clark (20 min)

"How Do Regulatory Sequences and Gene Regulation Vary Among Species, and How Can We Exploit this Variation?"
Mike Eisen (15 min)

"Human Genetics of Variation in Gene Expression"
Vivien Cheung (15 min)

"Synthetic Strategies for the Analysis of Combinatorial Cis-Regulation: From Yeast to Humans"
Barack Cohen (15 min)

2:45 p.m. - 3:00 p.m. ***Break***

3:00 p.m. - 4:15 p.m.

4c. Gene Regulatory Networks

Developing and validating network models; predictive models; transcriptional wiring diagram. Beyond individual cell states to dynamics (e.g., perturbations, developmental biology). Characterizing and understanding tissue-specific expression. What is the most important/meaningful/useful type of information to obtain in characterizing a gene regulatory network? What are the most promising experimental and analytical approaches, and how would they inform an understanding of the language of gene expression? For this topic in particular, what are the technical (experimental, computational, analytical) challenges to progress?

"The Explanatory Power of Animal Gene Regulatory Networks and the Challenges Ahead for Network Bioscience"
Chair: Eric Davidson (20 min)

"Dissecting the Logic of Cis-Regulatory Modules: Important Nodes Within Gene Regulatory Networks"
Angelika Stathopoulos (15 min)

"Unbiased Reconstruction of a Mammalian Transcriptional Network Mediating the Differential Response to Pathogens"
Aviv Regev (15 min)

"Molecular Sensors and Imaging for GRN Analysis at the Single-Cell Limit"
Scott Fraser (15 min)

4:30 p.m. - 5:30 p.m.

5. Overall Discussion Moderators: Rick Myers, Barbara Wold

Restatement of meeting objectives. What major ideas were raised? Are there conspicuous gaps that need to be discussed? How do these relate to understanding the regulatory code? Charge to breakout sessions.

5:30 p.m. - 6:15 p.m. ***Evening Break/Pick Up Dinner (45 min)***

6:15 p.m. - 7:45 p.m.

6. Breakout sessions follow topics 4a-c, with same chairs.
(See breakout assignment sheet.)

Deliverables:

- Major unanswered questions/challenges and their significance. What are the most important data sets, technologies, or basic questions, etc.?
- Which are most directed at understanding the regulatory code? How would the information inform understanding the regulatory code?
- How do they integrate with the other topics?
- How could they relate to the clinic (at least diagnosis, prognosis)?
- Of those, which are amenable to comprehensive approaches in 5 years? 10 years? Outline a viable pathway to getting there. Identify the scientific and technical obstacles.

8:00 p.m.

6. Breakout session chairs work on separate reports for next day.

Day 2, October 28, 2009

8:30 a.m. - 9:30 a.m.

7. Reports From Breakout Groups Moderator: Adam Felsenfeld

9:30 a.m. - 11:30 a.m.

8. Summary and Discussion Moderators: Peggy Farnham, Paul Sternberg

- How do these topics integrate?
- Prioritize the breakout proposals: “Top 10 list” of questions to be answered, with attendant challenges and technologies, etc.
- Examples of “genomic” projects that could be productive in providing the information that would help the community address the major questions.

Genomics of Gene Regulation Planning Workshop

Gaithersburg Marriott Washingtonian Center
October 27-28, 2009

PARTICIPANTS

Assya Abdallah
Scientific Program Analyst
Division of Extramural Research
National Human Genome Research Institute
National Institutes of Health
Suite 4076
5635 Fishers Lane
Rockville, MD 20892
(301) 496-7531
abdallahaf@mail.nih.gov

Karen Adelman, Ph.D.
Principal Investigator
Laboratory of Molecular Carcinogenesis
National Institute of Environmental Health Sciences
National Institutes of Health
Room D425
Mail Drop D4-02
111 Alexander Drive
Research Triangle Park, NC 27709
(919) 541-0001
adelmank@niehs.nih.gov

Richard A. Anderson, M.D., Ph.D.
Senior Scientific Officer
Division of Genetics and Developmental Biology
National Institute of General Medical Sciences
National Institutes of Health
Building 45, Room 2AS-25B
45 Center Drive, MSC 6200
Bethesda, MD 20892-6200
(301) 594-0943
andersor@nigms.nih.gov

David Arnosti, Ph.D.
Professor
Department of Biochemistry and Molecular Biology
Michigan State University
Biochemistry Building, Room 413
East Lansing, MI 48824-1319
(517) 432-5504
arnosti@msu.edu

Thomas S. Becker, Ph.D.
Chair of Developmental Neurobiology
Laboratory of Neurobiology and Genomics
Brain and Mind Research Institute
Sydney Medical School
University of Sydney
100 Mallett Street
Camperdown, NSW 2050
Australia
93510938
tsbecker@med.usyd.edu.au

Bradley E. Bernstein, M.D., Ph.D.
Associate Professor of Pathology
Department of Pathology
Harvard Medical School
Associate Pathologist
Massachusetts General Hospital
Associate Member
Broad Institute of MIT and Harvard
Early Career Scientist
Howard Hughes Medical Institute
Simches Research Building
CPZN 8234
185 Cambridge Street
Boston, MA 02114
(617) 726-6906
bernstein.bradley@mgh.harvard.edu

Mark D. Biggin, Ph.D.
Staff Scientist
Genomics Division
Lawrence Berkeley National Laboratory
MS 84-171
1 Cyclotron Road
Berkeley, CA 94720
(510) 486-7606
mdbiggin@lbl.gov

Paul Brazhnik, Ph.D.
Program Director
Center for Bioinformatics and Computational Biology
National Institute of General Medical Sciences
National Institutes of Health
MSC 6200
45 Center Drive
Bethesda, MD 20892-6200
(301) 451-6446
brazhnikp@mail.nih.gov

Michael Brodsky, Ph.D.
Assistant Professor
Program in Gene Function and Program in Molecular Medicine
University of Massachusetts Medical School
Lazare Research Building, Room 623
364 Plantation Street
Worcester, MA 01605
(508) 856-1640
michael.brodsky@umassmed.edu

Lisa Brooks, Ph.D.
Program Director
Genetic Variation Program
National Human Genome Research Institute
National Institutes of Health
5635 Fishers Lane
Bethesda, MD 20814
(301) 435-5544
lisa.brooks@nih.gov

Martha L. Bulyk, Ph.D.
Associate Professor of Medicine and Pathology and Health Sciences and Technology
Brigham and Women's Hospital
Harvard Medical School
New Research Building, Room 466D
77 Avenue Louis Pasteur
Boston, MA 02115
(617) 525-4725
mlbulyk@receptor.med.harvard.edu

Chris Burge, Ph.D.
Associate Professor
Department of Biology
Massachusetts Institute of Technology
68-223A
31 Ames Street
Cambridge, MA 02142
(617) 258-5997
cburge@mit.edu

Shawn Burgess, Ph.D.
Head
Developmental Genomics Section
Genome Technology Branch
National Human Genome Research Institute
National Institutes of Health
Building 50, Room 5537
50 South Drive
Bethesda, MD 20892
(301) 594-8224
burgess@mail.nih.gov

Anthony D. Carter, Ph.D.
Program Director
Division of Genetics and Developmental Biology
National Institute of General Medical Sciences
National Institutes of Health
MSC 6200
45 Center Drive
Bethesda, MD 20892-6200
(301) 594-0943
cartera@nigms.nih.gov

Vivian G. Cheung, M.D.
Investigator
Howard Hughes Medical Institute
Professor
Departments of Pediatrics and Genetics
W.W. Smith Chair in Pediatric Genomic Research
The Children's Hospital of Philadelphia
Department of Pediatrics
University of Pennsylvania
Abramson Research Center, Room 716B
3615 Civic Center Boulevard
Philadelphia, PA 19104
(215) 590-4950
vcheung@mail.med.upenn.edu

Andrew Clark, Ph.D.
Department of Molecular Biology and Genetics
Cornell University
Biotech Building, Room 107
Ithaca, NY 14853
(607) 255-0527
ac347@cornell.edu

Barak Alon Cohen, Ph.D.
Associate Professor of Genetics
Department of Genetics
Center for Genome Science
Washington University in St. Louis
Campus Box 8510
4444 Forest Park Parkway
St. Louis, MO 63106
(314) 362-3674
cohen@genetics.wustl.edu

Christine Colvis, Ph.D.
Director, Program Integration
National Institute on Drug Abuse
National Institutes of Health
6001 Executive Boulevard
Bethesda, MD 20892
(301) 443-6480
ccolvis@nida.nih.gov

James Niles Coulombe, Ph.D.
Program Officer
Developmental Biology, Genetics, and Teratology Branch
Center for Developmental Biology and Perinatal Medicine
Eunice Kennedy Shriver National Institute of Child Health and Human Development
National Institutes of Health
Building 6100, Room 4B01
MSC 7510
6100 Executive Boulevard
Bethesda, MD 20814-7510
(301) 451-1390
coulombej@mail.nih.gov

Robert Darnell, M.D., Ph.D.
Investigator and Professor
Howard Hughes Medical Institute
The Rockefeller University
Box 226
1230 York Avenue
New York, NY 10065
(212) 327-7474
smithje@rockefeller.edu

Eric H. Davidson, Ph.D.
Norman Chandler Professor of Cell Biology
California Institute of Technology
MC 156-29
1200 East California Boulevard
Pasadena, CA 91125
(626) 395-4937
davidson@caltech.edu

Emmanouil Dermitzakis, Ph.D.
Louis-Jeantet Professor
Department of Genetic Medicine and Development
University of Geneva Medical School
1 Rue Michel-Servet
Geneva 1211
Switzerland
3795464
emmanouil.dermitzakis@unige.ch

Michael Eisen, Ph.D., A.B.
Associate Professor of Genetics Genomics and Development
HHMI Investigator
Lawrence Berkeley National Lab
Howard Hughes Medical Institute
University of California Berkeley
UC Berkeley MC #3220
304B Stanley Hall
Berkeley, CA 94720
(510) 666-3639
mbeisen@lbl.gov

Laura Elnitski, Ph.D.
Investigator
Head
Genomic Functional Analysis Section
Genome Technology Branch
National Human Genome Research Institute
National Institutes of Health
5625 Fishers Lane
Rockville, MD 20878
(301) 451-0265
elnitski@mail.nih.gov

James Douglas Engel, Ph.D.
G. Carl Huber Professor and Chair
Department of Cell and Developmental Biology
University of Michigan Medical School
Biomedical Science Research Building, Room 3071
109 Zina Pitcher Place
Ann Arbor, MI 48109
(734) 615-7509
engel@umich.edu

Peggy J. Farnham, Ph.D.
Professor
Department of Pharmacology
Associate Director of Genomics
Genome Center
University of California, Davis
1 Shields Avenue
Davis, CA 95616
(530) 754-4988
pjfarnham@ucdavis.edu

Elise Feingold, Ph.D.
Program Director
Division of Extramural Research
National Human Genome Research Institute
National Institutes of Health
Suite 4076
5635 Fishers Lane
Rockville, MD 20852
(301) 496-7531
feingole@mail.nih.gov

Adam Felsenfeld, Ph.D.
Program Director
Division of Extramural Research
National Human Genome Research Institute
National Institutes of Health
Suite 4076
5635 Fishers Lane
Rockville, MD 20852
(301) 496-7531
felsenfa@mail.nih.gov

Gary Felsenfeld, Ph.D.
Chief
Physical Chemistry Section
NIH Distinguished Investigator
Laboratory of Molecular Biology
National Institute of Diabetes and Digestive and Kidney Diseases
National Institutes of Health
Building 5, Room 212
MSC 0540
9000 Rockville Pike
Bethesda, MD 20892-0540
(301) 496-4173
gary.felsenfeld@nih.gov

Scott E. Fraser, Ph.D.
Anna L. Rosen Professor of Biology and Engineering and Applied Science
Director
Biological Imaging Center
Director
Rosen Center for Biological Engineering
California Institute of Technology
MC 139-74
1200 East California Boulevard
Pasadena, CA 91125
(626) 395-2790
sefraser@caltech.edu

David K. Gifford , Ph.D.
Professor
MIT, Broad Institute, Whitehead Institute
32-G542 MIT Stata Center, 32 Vassar Street
Cambridge, MA 02493
(617) 253-6039
gifford@mit.edu

Peter Good, Ph.D.
Program Director
Division of Extramural Research
National Human Genome Research Institute
National Institutes of Health
Suite 4076
5635 Fishers Lane
Rockville, MD 20852
(301) 496-7531
goodp@mail.nih.gov

Eric Green, M.D., Ph.D.
Scientific Director
Genome Technology Branch
National Human Genome Research Institute
National Institutes of Health
Building 50, Room 5222
50 South Drive
Bethesda, MD 20892
(301) 402-2023
egreen@nhgri.nih.gov

Michael R. Green, M.D., Ph.D.
Investigator
Howard Hughes Medical Institute
Program in Gene Function and Expression
University of Massachusetts
Room 627
364 Plantation Street
Worcester, MA 01605
(508) 856-5330
michael.green@umassmed.edu

Alan Edward Guttmacher, M.D.
Acting Director
National Human Genome Research Institute
National Institutes of Health
9000 Rockville Pike
Bethesda, MD 20892
(301) 496-0844
guttmach@mail.nih.gov

Mark S. Guyer, Ph.D.
Director
Division of Extramural Research
National Genome Research Institute
National Institutes of Health
Suite 4123
5635 Fishers Lane
Rockville, MD 20892
(301) 496-7531
guyerm@exchange.nih.gov

Gordon Hager, Ph.D.
Chief
Lab of Receptor Biology and Gene Expression
Chair
Center of Excellence in Chromosome Biology
National Cancer Institute
National Institutes of Health
Building 41, Room B602
Bethesda, MD 20892
(301) 496-9867
hagerg@exchange.nih.gov

Ross C. Hardison, Ph.D.
T. Ming Chu Professor of Biochemistry and Molecular Biology
Center for Comparative Genomics and Bioinformatics
Department of Biochemistry and Molecular Biology
The Pennsylvania State University
Wartik Laboratory, Room 304
University Park, PA 16802
(814) 863-0113
rch8@psu.edu

Mark Johnston, Ph.D.
Department of Biochemistry and Molecular Genetics
University of Colorado Denver
Anschutz Medical Campus
Mail Stop 8101
P.O. Box 6511
Aurora, CO 80045
(303) 724-3201
mark.johnston@udenver.edu

Jack D. Keene, Ph.D.
Professor of Molecular Genetics and Microbiology
Duke University Medical Center
P.O. Box 3020
Durham, NC 27710
(919) 684-5138
keene001@mc.duke.edu

Manolis Kellis, Ph.D., M.Eng.
Associate Professor of Computer Science
Broad Institute of MIT and Harvard
Room 32G-842
32 Vassar Street
Cambridge, MA 02139
(617) 253-2419
manoli@mit.edu

Stuart Kim, Ph.D.
Professor
Department of Developmental Biology
Stanford University Medical Center
Stanford, CA 94305
(650) 725-7671
stuartkm@stanford.edu

Robert E. Kingston, Ph.D.
Chief
Department of Molecular Biology
Massachusetts General Hospital
CPZN 7250
185 Cambridge Street
Boston, MA 02114
(617) 726-5990
kingston@molbio.mgh.harvard.edu

Donna Krasnewich, M.D., Ph.D.
Program Director
Division of Genetics and Developmental Biology
National Institute of General Medical Sciences
National Institutes of Health
Building 45, Room 2AS25H
Bethesda, MD 20892
(301) 594-0943
dkras@mail.nih.gov

Leonid Kruglyak, Ph.D.
Professor of Ecology and Evolutionary Biology
Chair
Graduate Program in Quantitative and Computational Biology, Ecology, and Evolutionary Biology
Lewis-Sigler Institute for Integrative Genomics
Princeton University/Howard Hughes Medical Institute
Carl Icahn Laboratory, Room 143
Princeton, NJ 08544
(609) 258-9209
leonid@genomics.princeton.edu

Robb E. Krumlauf, Ph.D.
Scientific Director
Stowers Institute for Medical Research
Professor
Neuroscience Graduate Program
Professor
Department of Anatomy and Cell Biology
Professor of Oral Biology
School of Dentistry
Stowers Institute for Medical Research
University of Kansas
1000 East 50th Street
Kansas City, MO 64110
(816) 926-4051
rek@stowers.org

Patrick Lemaire, Ph.D., M.S.
Institut de Biologie du Development de Marseille-Luminy
Centre National de la Recherche Scientifique
Université de la Méditerranée
Campus de Luminy, Case 907
Marseille cedex 9 F-13288
France
63807
lemaire@ibdml.univ-mrs.fr

Boris Lenhard, Ph.D.
Computational Biology Unit
Bergen Center for Computational Science
Sars Centre for Marine Molecular Biology
University of Bergen
Thormohlensgate 55
Bergen 5008
Norway
4.797e+09
boris.lenhard@bccs.uib.no

Michael S. Levine, Ph.D.
Head and Professor
Division of Genetics, Genomics, and Development
Department of Mol Cell Biology
University of California at Berkeley
Berkeley, CA 94720
(510) 642-5014
mlevine@berkeley.edu

Jason Lieb, Ph.D.
Associate Professor
Department of Biology and Carolina Center for Genome Sciences
The University of North Carolina at Chapel Hill
Campus Box 3280
Fordham Hall, Room 408
Chapel Hill, NC 27599-3280
(919) 843-3228
jlieb@bio.unc.edu

John T. Lis, Ph.D.
Barbara McClintock Professor of Molecular Biology and Genetics
Department of Molecular Biology and Genetics
Cornell University
Biotechnology Building
Ithaca, NY 14853
(607) 255-2442
jtl10@cornell.edu

Rebecca Lowdon
Scientific Program Analyst
National Human Genome Research Institute
National Institutes of Health
Apartment 415
3220 Connecticut Avenue
Washington, DC 20008
(540) 580-8816
lowdonrf@mail.nih.gov

Paul Meltzer, M.D., Ph.D.
Chief
Genetics Branch
Center for Cancer Research
National Cancer Institute
National Institutes of Health
37 Convent Drive
Bethesda, MD 20892
(301) 496-5266
pmeltzer@mail.nih.gov

Alan Mark Michelson, M.D., Ph.D.
Director
Genetics and Developmental Biology Center
Division of Intramural Research
National Heart, Lung, and Blood Institute
National Institutes of Health
Building 10, Room 6C-103A
Bethesda, MD 20892
(301) 451-8041
michelsonam@nhlbi.nih.gov

Judy Mietz, Ph.D., M.S.
Program Director and Chief
DNA and Chromosome Aberrations Branch
Division of Cancer Biology
National Cancer Institute
National Institutes of Health
Executive Plaza North, Room 5028
6130 Executive Boulevard
Rockville, MD 20852
(301) 496-9326
mietzj@mail.nih.gov

Leonid Mirny, Ph.D.
Massachusetts Institute of Technology
Room E25-526
77 Massachusetts Avenue
Cambridge, MA 02139
(617) 452-4862
leonid@mit.edu

Richard Myers, Ph.D.
Director and President
HudsonAlpha Institute for Biotechnology
601 Genome Way
Huntsville, AL 35806
(256) 327-0431
rmyers@hudsonalpha.org

Tomi Pastinen, M.D., Ph.D.
Assistant Professor
Canada Research Chair in Human Genomics
Departments of Human and Medical Genetics
Genome Quebec Innovation Centre
McGill University
Room 6202
740 Dr. Penfield Avenue
Montreal H3A 1A4
Quebec
Canada
(514) 398-1777
tomi.pastinen@mcgill.ca

William Pavan, Ph.D.
Senior Investigator
Genetic Disease Research Branch
National Human Genome Research Institute
National Institutes of Health
Building 49, Room 4A-82
49 Convent Drive
Bethesda, MD 20892
(301) 496-7584
bpavan@mail.nih.gov

Len Pennacchio, Ph.D.
Senior Staff Scientist
Genomics Division
Lawrence Berkeley National Laboratory and Joint Genome Institute
Genomics Division, Room B84
MS 84-171
1 Cyclotron Road
Berkeley, CA 94720
(925) 980-3711
lapennacchio@lbl.gov

Jane Peterson, Ph.D.
Associate Director
Division of Extramural Research
National Human Genome Research Institute
National Institutes of Health
Suite 4076
5635 Fishers Lane
Rockville, MD 20852
(301) 496-7531
petersoj@mail.nih.gov

Chris Ponting, Ph.D., M.S.
Professor of Genomics and MRC Program Leader
Functional Genomics Unit
Medical Research Council and University of Oxford
Department of Physiology, Anatomy, and Genetics
Le Gros Clark Building
South Parks Road
Oxford OX1 3QX
United Kingdom
+44 1865 285855
chris.ponting@dpag.ox.ac.uk

Jonathan Pritchard, Ph.D.
Professor
Department of Human Genetics
University of Chicago
Cummings Life Science Center, Room 507
920 East 58th Street
Chicago, IL 60637
(773) 834-0525
vwebster@bsd.uchicago.edu

Aviv Regev, Ph.D.
Early Career Scientist
Howard Hughes Medical Institute
Core Member
Broad Institute of MIT and Harvard
Assistant Professor of Biology
Massachusetts Institute of Technology
7 Cambridge Center
Cambridge, MA 02142
(617) 714-7020
aregev@broadinstitute.org

Anna Rossoshek, M.B.A., M.S.
Senior Scientific Analyst
Division of Extramural Research
National Human Genome Research Institute
National Institutes of Health
Suite 4076
5635 Fishers Lane
Rockville, MD 20852
(301) 451-8323
rossosheka@mail.nih.gov

John Satterlee, Ph.D.
Program Director
Roadmap Epigenomics Program Co-Coordinator
National Institute on Drug Abuse
National Institutes of Health
Room 4267
MSC 9555
6001 Executive Boulevard
Bethesda, MD 20892-9555
(301) 435-1020
satterleej@nida.nih.gov

Jeffery A. Schloss, Ph.D.
Program Director, Technology Development Coordination
Division of Extramural Research
National Human Genome Research Institute
National Institutes of Health
MSC 9305
5635 Fishers Lane
Bethesda, MD 20892-9305
(301) 496-7531
jeff_schloss@nih.gov

Arend Sidow, Ph.D.
Departments of Pathology and Genetics
Stanford University
Stanford University Medical Center, Room R353
300 Pasteur Drive
Stanford, CA 94305
(650) 498-7024
arend@stanford.edu

Eric Siggia, Ph.D.
The Rockefeller University
1230 York Avenue
New York, NY 10065
(212) 327-8546
siggiae@rockefeller.edu

Angela Stathopoulos, Ph.D.
Assistant Professor of Biology
Division of Biology
California Institute of Technology
1200 East California Boulevard
MC 114-96
Pasadena, CA 91125
(626) 395-5855
angelike@caltech.edu

Paul W. Sternberg, Ph.D.
Thomas Hunt Morgan Professor of Biology
Division of Biology
California Institute of Technology
Investigator
Howard Hughes Medical Institute
Room 156-29
1200 East California Boulevard
Pasadena, CA 91125
(626) 395-2181
pws@caltech.edu

Gary D. Stormo, Ph.D.
Department of Genetics
Center for Genome Sciences
School of Medicine
Washington University at St. Louis
Box 8510
4444 Forest Park Boulevard
St. Louis, MO 63108
(314) 747-5534
stormo@genetics.wustl.edu

Jeffery Struewing, M.D., M.S.
Program Director
National Human Genome Research Institute
National Institutes of Health
Suite 4076
5635 Fishers Lane
Rockville, MD 20852
301-594-3757
struewij@mail.nih.gov

Kevin Struhl, Ph.D., M.S.
David Wesley Gaiser Professor of Biological Chemistry and Molecular Pharmacology
Harvard Medical School
240 Longwood Avenue
Boston, MA 02115
617 432-2104
kevin@hms.harvard.edu

Jason Swedlow, Ph.D.
Wellcome Trust Centre for Gene Regulation and Expression
College of Life Sciences
University of Dundee
Dow Street
Dundee, Scotland
United Kingdom
1.382e+09
jason@lifesci.dundee.ac.uk

Danilo Angeles Tagle, Ph.D.
Program Director in Neurogenetics
National Institute of Neurological Disorders and Stroke
National Institutes of Health
Neuroscience Building, Room 2114
6001 Executive Boulevard
Bethesda, MD 20892
(301) 496-5745
tagled@ninds.nih.gov

Gary F. Temple, M.D.
Program Director
Division of Extramural Research
National Human Genome Research Institute
National Institutes of Health
Suite 4079
MSC 9305
5635 Fishers Lane
Bethesda, MD 20892-9305
(301) 594-5951
gtemple@mail.nih.gov

Laurie Tompkins, Ph.D.
Chief
Genetic Mechanisms Branch
Division of Genetics and Developmental Biology
National Institute of General Medical Sciences
National Institutes of Health
MSC 6200
45 Center Drive
Bethesda, MD 20892-6200
(301) 594-0943
tompkinL@nigms.nih.gov

Marc Vidal, Ph.D.
Professor of Genetics
Cancer Biology
Dana-Farber Cancer Institute
Harvard Medical School
SM 858
44 Binney Street
Boston, MA 02115
(617) 632-5180
marc_vidal@dfci.harvard.edu

Lu Wang, Ph.D.
Program Director
Division of Extramural Research
National Institutes of Health
Suite 4105
5635 Fishers Lane
MSC 9305
Bethesda, MD 20892
(301) 594-7303
wanglu@mail.nih.gov

Robert Waterston, M.D., Ph.D.
William H. Gates III Endowed Chair in
Biomedical Sciences
National Human Genome Research Institute
University of Washington
Box 355065
1705 NE Pacific Street
Seattle, WA 98195-5065
(206) 685-7347
waterston@gs.washington.edu

Barbara Wold, Ph.D.
Biology Division
California Institute of Technology
1200 East California Boulevard
Pasadena, CA 91125
(626) 395-4916
woldb@caltech.edu

Richard A. Young, Ph.D.
Whitehead Institute for Biomedical
Research
Massachusetts Institute of Technology
9 Cambridge Center
Cambridge, MA 02142
(617) 258-5219
young@wi.mit.edu

Keji Zhao
Senior Investigator
National Heart, Lung, and Blood Institute
National Institutes of Health
9000 Rockville Pike
Bethesda, MD 20892
(301) 496-2098
zhaok@nhlbi.nih.gov