Recommendations related to genome function
from NHGRI's Planning Workshop on the Future Opportunities for Genome Sequencing & Beyond

Mark Gerstein & Rick Myers

Recommendations related to genome function

Part 1: Workshop summary (Mark)

Part 2: Our Additional Suggestions (Rick)

- Overall Key Message of both Parts
  - Functional annotation is **critical** for NHGRI's mission
  - Biotechnological & computational **tools** need to be developed for this
  - Need a **systematic catalog** of non-coding elements & their interactions

Recommendations related to genome function

Part 1: Workshop summary (Mark)

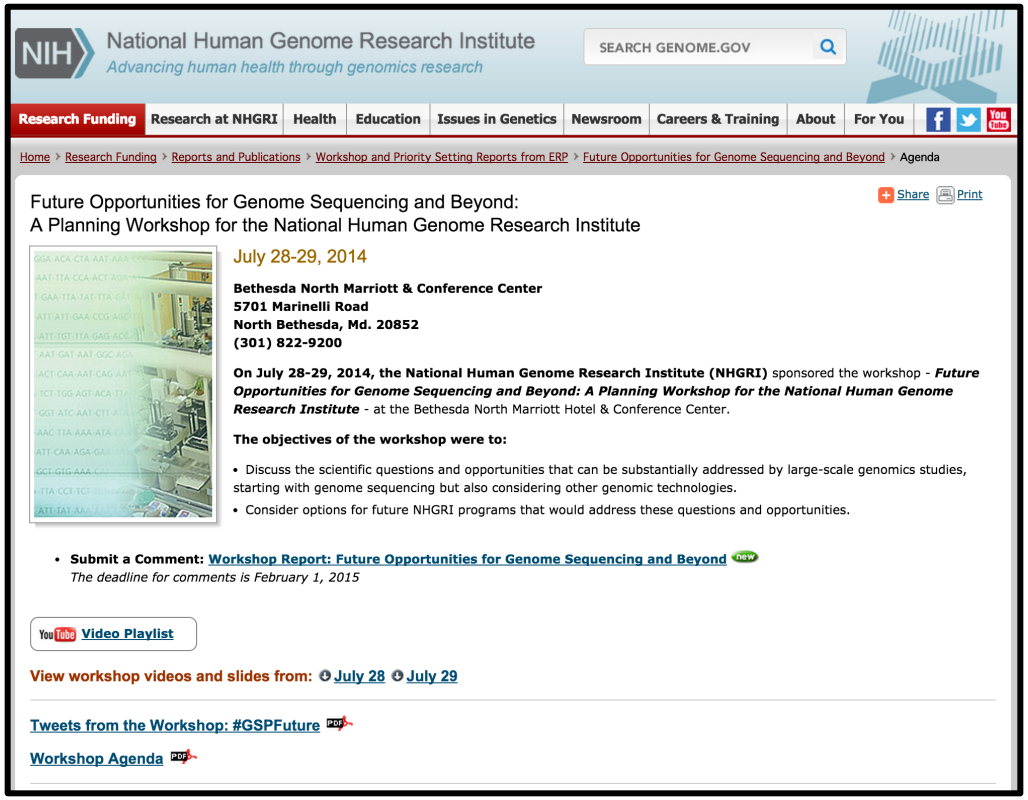
Part 2: Our Additional Suggestions (Rick)

- Overall Key Message of both Parts
  - Functional annotation is **critical** for NHGRI's mission
  - Biotechnological & computational **tools** need to be developed for this
  - Need a **systematic catalog** of non-coding elements & their interactions

Workshop last summer: Future Opportunities for Genome Sequencing & Beyond

Goals:

- "Discuss Qs & opportunities addressed using genomics studies, starting with genomic sequencing but also using other technologies"
- "Consider future NHGRI programs addressing" these



Slides and videos available at

Genome.gov/27558042

Backdrop: A Future where N**HG**RI is only involved in minority of **HG** sequencing

- **Human Genome Sequencing is a success story… but what to do now?**
- Most genome sequencing will be done outside of NHGRI
- "NHGRI needs to position itself to positively influence the large amount of sequencing that will occur."
- Involving itself in **partnerships**
- Creating **exportable technology, platforms & standards** for interpreting the genome
- Focusing on **things that scale**

Sections within Workshop

- **Genomic Architecture of Disease at Scale**
- **Integrating Genomic Variant Discovery with Function**
- **Clinical Genomic Sequencing at Scale**
- **Comparative & Evolutionary Genomics**

[Text adapted from workshop report genome.gov/27559219]

Sections within Workshop

- **Genomic Architecture of Disease at Scale**
  - Exec Summary : "Using genomic sequencing to determine variants underlying human disease & healthy traits, including both Mendelian & complex diseases, **continues to be an important activity** & will need to be addressed at scale."
- **Integrating Genomic Variant Discovery with Function**
  - "Improving our understanding of the impact of variants through functional genomics studies **is critical** to inform gene-disease relationships."
- **Clinical Genomic Sequencing at Scale**
  - "Translating genomics to medical practice **will require a critical evaluation** of the utility of sequencing & approaches to clinical implementation."
- **Comparative & Evolutionary Genomics**
  - "This **is still needed** to inform the prioritization and interpretation of genetic variants."

[Text adapted from workshop report genome.gov/27559219]

Integrating Genomic Variant Discovery with Function

- **1** Overview talk by J Ecker
- **Discussion Group** (**~25**):
  Mark Adams, Toby Bloom, Jim Broach, Carol Bult, Carlos Bustamante, Deborah Colantuoni, Joe Ecker, Elise Feingold, Kelly Frazer, Ross Hardison, Chanita Hughes-Halbert, Stephen Kingsmore, Jim Lupski, Gabor Marth, Debbie Nickerson, Mike Pazin, Len Pennacchio, Ulrike Peters, Aviv Regev, Jay Shendure, Mike Snyder, Simona Volpi, Peter Good
- Discussion summary by **2** (R Myers & M Gerstein)

Opportune time to study "function" on a large scale

- Huge number of variants available from many studies from NHGRI & beyond
  - Functional characterization = connection between genomes & biology
- Recent development of new technologies
  - CRISPR, large-scale epigenomics, single cell, etc.

We need a foundational resource to integrate functional information on many discovered variants

What should the resource be?

- Different types of function
  - At molecular/biochemical and cellular levels
    - can be studied at scale & systematized
    - Also, is closer to the variants
  - At organismal level
    - Not as easy to scale or to systematize
- NHGRI should find the "sweet spot"
  - Problems that capitalize on the new technologies
    - Lots of readout with modest investment
  - Best models – cells? mouse? model diseases?

Dichotomy of Directions

- Top-down: Develop catalogs of elements &/ or all possible variants & then intersect them with variants found in disease studies
  - ex: Shendure challenge talk

- Bottom-up: Start from a list of disease variants & characterize them functionally

Both have merits

Multiple Approaches

- Approaches that look at large numbers of genes, variants, cell types, etc. in a standardized, high-throughput way

- In contrast: Deep disease/gene-specific studies
  - Require domain experts & detailed assays, many of which cannot be scaled
  - Not the province of NHGRI -- at least not on their own

- Important to have both & integrate them
  - Build special informatics infrastructure to tie them together

Other considerations

- Scaling from the genome-scale assays to population-scale
  - Success of eQTL & related projects
  - Personal functional genomics, value in longitudinal studies

- Functional genomics is valuable beyond just variant characterization
  - Use high-throughput sequencing to characterize cell types
    - e.g., to develop cellular biomarkers
    - ex: Regev challenge talk (Single-cell transcriptomics & Human Cell Atlas Project)

NHGRI synthesis of broader discussion (involving **>100** people) of goals, opportunities & recommendations from the meeting

Rec #1: Defining the function of coding & noncoding sequences is foundational for genomics

- Develop & deploy assays reporting disease-relevant functions at the variant, gene & **pathway levels**

[Text adapted from workshop report genome.gov/27559219]

Rec #1: Defining the function of coding & noncoding sequences is foundational for genomics

- Develop & deploy assays reporting disease-relevant functions at the variant, gene & **pathway levels**
- Functions should be considered at different scales (e.g. molecular v. cellular)
- NHGRI should consider both function-first approaches and variant-first approaches.
- **Computational methods need to be developed** to predict the effect of coding and non-coding variants

[Text adapted from workshop report genome.gov/27559219]

Rec #2: Develop tools to manipulate genomic sequences at scale

& experimentally characterize their impact

- **Need develop new ways to measure function & determine causality of stat. significant variants**
- **NHGRI should raise the technical challenge on how to scale up (to whole genome or whole population) the most important functional assays**, while maintaining assay validity.

[Text adapted from workshop report genome.gov/27559219]

Rec #2: Develop tools to manipulate genomic sequences at scale

& experimentally characterize their impact

- **Need develop new ways to measure function & determine causality of stat. significant variants**
- **NHGRI should raise the technical challenge on how to scale up (to whole genome or whole population) the most important functional assays**, while maintaining assay validity.
- Large scale assays need to be scaled from molecular to organ, organism, and clinical levels.
- Need to improve understanding on how proteins interact with the genome.
- NHGRI could help foster assays and models that allow us to test how drugs & other environmental agents interact with the genome.
- Personal genomics can be expanded to include personal functional genomics.

[Text adapted from workshop report genome.gov/27559219]

Rec #3: Need to systematically catalog molecular components & their interactions

- **Function should also be considered at the systems biology level**
- The catalog of regulatory elements is not complete. Additional profiling of regulatory data needs to be done in key tissues and cell types.

[Text adapted from workshop report genome.gov/27559219]

Rec #3: Need to systematically catalog molecular components & their interactions

- **Function should also be considered at the systems biology level**
- The catalog of regulatory elements is not complete. Additional profiling of regulatory data needs to be done in key tissues and cell types.
- Functional genomics is valuable beyond characterization of variants. Sequencing can be used to characterize cell types.
- **NHGRI should probably limit consideration to genetic effects, and should not further consider environmental…or microbiome effects**

[Text adapted from workshop report genome.gov/27559219]

Recommendations related to genome function

Part 1: Workshop summary (Mark)

Part 2: Our Additional Suggestions (Rick)

- Overall Key Message of both Parts
  - Functional annotation is **critical** for NHGRI's mission
  - Biotechnological & computational **tools** need to be developed for this
  - Need a **systematic catalog** of non-coding elements & their interactions

Some important, crosscutting themes

Large-scale genomics projects have helped to establish:

- Standardized experimental and analytical approaches
- Quality control metrics, resource development, economies of scale, and data sharing
- Generation of multiple data types and large studies that lead to meta-analysis and integration

Some important, crosscutting themes

Large-scale genomics projects have helped to establish:

- Standardized experimental and analytical approaches
- Quality control metrics, resource development, economies of scale, and data sharing
- Generation of multiple data types and large studies that lead to meta-analysis and integration

Small- and medium-scale genomics projects have advantages:

- More detailed knowledge and interest in specific biological problems
- Often geared towards determining mechanism, testing specific hypotheses
- Support wide range of researchers

Some important, crosscutting themes

Knowledge of genome function is essential to understand human biology and genomic basis of disease

Some important, crosscutting themes

Knowledge of genome function is essential to understand human biology and genomic basis of disease

Evolutionary genomics is important

- Sequencing and alignment of primate and vertebrate genomes has helped identify >3 million conserved elements

Some important, crosscutting themes

Knowledge of genome function is essential to understand human biology and genomic basis of disease

Evolutionary genomics is important

- Sequencing and alignment of primate and vertebrate genomes has helped identify >3 million conserved elements

We have learned an incredible amount by identifying human knock-outs (i.e. Mendelian diseases) and, to a smaller extent, dominant mutations

1: Additional suggestions for NHGRI

Don’t try to do everything in genomics: Partner with other NIH institutes

1: Additional suggestions for NHGRI

Don’t try to do everything in genomics: Partner with other NIH institutes

Many very large-scale whole genome sequencing and functional genomics DNA projects are already being done -- and will continue to be done -- by other institutes

- Disease/organ-specific genomics

1: Additional suggestions for NHGRI

Don't try to do everything in genomics: Partner with other NIH institutes

Many very large-scale whole genome sequencing and functional genomics DNA projects are already being done -- and will continue to be done -- by other institutes

Disease/organ-specific genomics

But NHGRI can and should still play a major role in these studies (as follows)

2: Additional suggestions for NHGRI

Continue to support development of new DNA sequencing technologies (and other high-throughput genomics/genetics approaches)

2: Additional suggestions for NHGRI

Continue to support development of new DNA sequencing technologies (and other high-throughput genomics/genetics approaches)

Support work that allows us to analyze every single base pair accurately, regardless of length or context of sequence variants

- e.g., missing sequence data may miss undiagnosed diseases

2: Additional suggestions for NHGRI

Continue to support development of new DNA sequencing technologies (and other high-throughput genomics/genetics approaches)

Support work that allows us to analyze every single base pair accurately, regardless of length or context of sequence variants

- e.g., missing sequence data may miss undiagnosed diseases

Emphasize phenotypes

2: Additional suggestions for NHGRI

Continue to support development of new DNA sequencing technologies (and other high-throughput genomics/genetics approaches)

Support work that allows us to analyze every single base pair accurately, regardless of length or context of sequence variants

- e.g., missing sequence data may miss undiagnosed diseases

Emphasize phenotypes

Give much more support for advances in bioinformatics, computing and data analysis

2: Additional suggestions for NHGRI

Continue to support development of new DNA sequencing technologies (and other high-throughput genomics/genetics approaches)

Support work that allows us to analyze every single base pair accurately, regardless of length or context of sequence variants

- e.g., missing sequence data may miss undiagnosed diseases

Emphasize phenotypes

Give much more support for advances in bioinformatics, computing and data analysis

Take a more systems biology approach

- Consider the effects of genetic variants in the larger context of the cell, the individual and the environment
- Perturbation approaches

3: Additional suggestions for NHGRI

Put a much heavier emphasis on integration functional studies with evolutionary information

This is not just measurement of conservation. Must integrate phylogeny, sequence similarity, human population genetics

Continue to improve predictions of deleterious mutations (e.g., CADD, tolerance scores, FunSeq, etc.)

4: Additional suggestions for NHGRI

Foster new and emerging methods for assessing function of base pairs in the genome

- Some production-type grants (but not too large)
- R01s to support creative technology development (perhaps partner with NIGMS)
- Very high-throughput genome editing
- Ultra-high throughput functional assays

5: Additional suggestions for NHGRI

Be the lead in cataloguing different sequencing efforts

5: Additional suggestions for NHGRI

Be the lead in cataloguing different sequencing efforts

Create data sharing standards for more interoperability and collaboration

5: Additional suggestions for NHGRI

Be the lead in cataloguing different sequencing efforts

Create data sharing standards for more interoperability and collaboration

Continue to expand beyond whole genome sequencing (e.g., very broad and deep functional genomics studies)

5: Additional suggestions for NHGRI

Be the lead in cataloguing different sequencing efforts

Create data sharing standards for more interoperability and collaboration

Continue to expand beyond whole genome sequencing (e.g., very broad and deep functional genomics studies)

Include large-scale efforts in epigenomics and metabolomics

5: Additional suggestions for NHGRI

Be the lead in cataloguing different sequencing efforts

Create data sharing standards for more interoperability and collaboration

Continue to expand beyond whole genome sequencing (e.g., very broad and deep functional genomics studies)

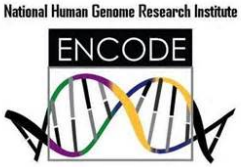
Include large-scale efforts in epigenomics and metabolomics

Continue to lead in training genome scientists, clinical researchers, and support diversity

Why coordination would be good:

#s of RNA-seq datasets from different consortia

ENCODE: 588



REMC: 166

GTEx: 837

TCGA: 24,662

ICGC: 1,486

And many other functional genomic datasets are available from different consortia

For example:

Establish a coordinated data integration and analysis hub

For example:

Establish a coordinated data integration and analysis hub

Integrate NHGRI-funded standardized data with large amount of other functional genomics data produced worldwide

For example:

Establish a coordinated data integration and analysis hub

Integrate NHGRI-funded standardized data with large amount of other functional genomics data produced worldwide

Make functional genomics data into digestible annotation that can be interrelated with a variety of variants (common, rare, SVs, disease, etc.)

For example:

Establish a coordinated data integration and analysis hub

Integrate NHGRI-funded standardized data with large amount of other functional genomics data produced worldwide

Make functional genomics data into digestible annotation that can be interrelated with a variety of variants (common, rare, SVs, disease, etc.)

Develop standardized and exportable pipelines, QC metrics, and analysis approaches

For example:

Establish a coordinated data integration and analysis hub

Integrate NHGRI-funded standardized data with large amount of other functional genomics data produced worldwide

Make functional genomics data into digestible annotation that can be interrelated with a variety of variants (common, rare, SVs, disease, etc.)

Develop standardized and exportable pipelines, QC metrics, and analysis approaches

Amazon model: build a shopping platform on a core (i.e., books) and then expand it outwards to everything

Recommendations related to genome function

Part 1: Workshop summary (Mark)

Part 2: Our Additional Suggestions (Rick)

- Overall Key Message of both Parts
  - Functional annotation is **critical** for NHGRI's mission
  - Biotechnological & computational **tools** need to be developed for this
  - Need a **systematic catalog** of non-coding elements & their interactions

Recommendations related to genome function

Part 1: Workshop summary (Mark)

Part 2: Our Additional Suggestions (Rick)

- Overall Key Message of both Parts
  - Functional annotation is **critical** for NHGRI's mission
  - Biotechnological & computational **tools** need to be developed for this
  - Need a **systematic catalog** of non-coding elements & their interactions